Merrill Skolnik

RADAR HANDBOOK

Third Edition

ABOUT THE EDITOR IN CHIEF

Merrill Skolnik was Superintendent of the Radar Division at the U.S. Naval Research Laboratory for over 30 years. Before that he was involved in advances in radar while at the MIT Lincoln Laboratory, the Institute for Defense Analyses, and the Research Division of Electronic Communications, Inc. He is the author of the popular McGraw-Hill textbook *Introduction to Radar Systems*, now in its third edition, the editor of *Radar Applications*, as well as being a former editor of the *Proceedings of the IEEE*. He earned the Doctor of Engineering Degree from The Johns Hopkins University, where he also received the B.E and M.S.E degrees in electrical engineering. He is a member of the U.S. National Academy of Engineering, a Fellow of the IEEE, and the first recipient of the IEEE Dennis J. Picard Medal for Radar Technologies and Applications.

RADAR HANDBOOK

Merrill I. Skolnik

Editor in Chief
Third Edition

New York Chicago San Francisco Lisbon London Madrid
Mexico City Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

The McGraw-Hill Companies

Cataloging-in-Publication Data is on file with the Library of Congress

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative, please visit the Contact Us pages at www.mhprofessional.com.

Radar Handbook, Third Edition

Copyright © 2008 by The McGraw-Hill Companies.

All rights reserved. Printed in the United States of America. Except as permitted under the Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of publisher.

1 2 3 4 5 6 7 8 9 0 DOC DOC 0 1 9 8

ISBN 978-0-07-148547-0
MHID 0-07-148547-3

Sponsoring Editor Wendy Rinaldi
Editorial Supervisor Janet Walden
Project Editor LeeAnn Pickrell
Acquisitions Coordinator Mandy Canales
Copy Editor LeeAnn Pickrell
Proofreader Susie Elkind
Production Supervisor Jean Bodeaux
Composition International Typesetting & Composition
Illustration International Typesetting & Composition
Art Director, Cover Jeff Weeks
Cover Designer Mary McKeon

Information has been obtained by McGraw-Hill from sources believed to be reliable. However, because of the possibility of human or mechanical error by our sources, McGraw-Hill, or others, McGraw-Hill does not guarantee the accuracy, adequacy, or completeness of any information and is not responsible for any errors or omissions or the results obtained from the use of such information.

CONTENTS

Contributors **xiii**

Preface **xv**

Chapter 1 An Introduction and Overview of Radar **1.1**
Merrill Skolnik

1.1 Radar in Brief / *1.1*

1.2 Types of Radars / *1.5*

1.3 Information Available from a Radar / *1.7*

1.4 The Radar Equation / *1.10*

1.5 Radar Frequency Letter-band Nomenclature / *1.13*

1.6 Effect of Operating Frequency on Radar / *1.14*

1.7 Radar Nomenclature / *1.18*

1.8 Some Past Advances in Radar / *1.19*

1.9 Applications of Radar / *1.20*

1.10 Conceptual Radar System Design / *1.22*

Chapter 2 MTI Radar **2.1**
William W. Shrader and Vilhelm Gregers-Hansen

2.1 Preface / *2.1*

2.2 Introduction to MTI Radar / *2.2*

2.3 Clutter Filter Response to Moving Targets / *2.9*

2.4 Clutter Characteristics / *2.10*

2.5 Definitions / *2.19*

2.6 Improvement Factor Calculations / *2.23*

2.7 Optimum Design of Clutter Filters / *2.25*

2.8 MTI Clutter Filter Design / *2.33*

2.9 MTI Filter Design for Weather Radars / *2.46*

2.10 Clutter Filter Bank Design / *2.52*

2.11 Performance Degradation Caused by Receiver Limiting / *2.59*

2.12 Radar System Stability Requirements / *2.65*

2.13 Dynamic Range and A/D Conversion Considerations / *2.78*

2.14 Adaptive MTI / *2.80*

2.15 Radar Clutter Maps / *2.83*

2.16 Sensitivity-velocity Control (SVC) / *2.87*
2.17 Considerations Applicable to MTI Radar Systems / *2.91*

Chapter 3 Airborne MTI **3.1**
James K. Day and Fred M. Staudaher
3.1 Systems Using Airborne MTI Techniques / *3.1*
3.2 Coverage Considerations / *3.2*
3.3 Airborne MTI Performance Drivers / *3.3*
3.4 Platform Motion and Altitude Effects on MTI Performance / *3.3*
3.5 Platform-motion Compensation Abeam / *3.10*
3.6 Scanning-motion Compensation / *3.14*
3.7 Simultaneous Platform Motion and Scan Compensation / *3.18*
3.8 Platform-motion Compensation, Forward Direction / *3.21*
3.9 Space-time Adaptive Motion Compensation / *3.23*
3.10 Effect of Multiple Spectra / *3.31*
3.11 Example AMTI Radar System / *3.32*

Chapter 4 Pulse Doppler Radar **4.1**
John P. Stralka and William G. Fedarko
4.1 Characteristics and Applications / *4.1*
4.2 Pulse Doppler Clutter / *4.14*
4.3 Dynamic-range and Stability Requirements / *4.24*
4.4 Range and Doppler Ambiguity Resoluton / *4.31*
4.5 Mode and Waveform Design / *4.35*
4.6 Range Performance / *4.39*
List of Abbreviations / *4.48*

Chapter 5 Multifunctional Radar Systems for Fighter Aircraft **5.1**
David Lynch, Jr. and Carlo Kopp
5.1 Introduction / *5.1*
5.2 Typical Missions and Modes / *5.10*
5.3 A-A Mode Descriptions & Waveforms / *5.16*
5.4 A-S Mode Descriptions & Waveforms / *5.28*

Chapter 6 Radar Receivers **6.1**
Michael E. Yeomans
6.1 The Configuration of a Radar Receiver / *6.1*
6.2 Noise and Dynamic-range Considerations / *6.4*
6.3 Bandwidth Considerations / *6.9*
6.4 Receiver Front End / *6.10*
6.5 Local Oscillators / *6.14*
6.6 Gain Control / *6.22*
6.7 Filtering / *6.24*

6.8 Limiters / *6.29*
6.9 I/Q Demodulators / *6.31*
6.10 Analog-to-Digital Converters / *6.35*
6.11 Digital Receivers / *6.40*
6.12 Diplex Operation / *6.46*
6.13 Waveform Generation and Upconversion / *6.47*

Chapter 7 Automatic Detection, Tracking, and Sensor Integration **7.1**
W. G. Bath and G. V.Trunk
7.1 Introduction / *7.1*
7.2 Automatic Detection / *7.1*
7.3 Automatic Tracking / *7.22*
7.4 Networked Radars / *7.46*
7.5 Unlike-sensor Integration / *7.49*

Chapter 8 Pulse Compression Radar **8.1**
Michael R. Ducoff and Byron W. Tietjen
8.1 Introduction / *8.1*
8.2 Pulse Compression Waveform Types / *8.2*
8.3 Factors Affecting Choice of Pulse Compression Systems / *8.26*
8.4 Pulse Compression Implementation and Radar System Examples / *8.28*
Appendix / *8.36*

Chapter 9 Tracking Radar **9.1**
Dean D. Howard
9.1 Introduction / *9.1*
9.2 Monopulse (Simultaneous Lobing) / *9.3*
9.3 Scanning and Lobing / *9.16*
9.4 Servosystems for Tracking Radar / *9.17*
9.5 Target Acquisition and Range Tracking / *9.20*
9.6 Special Monopulse Techniques / *9.24*
9.7 Sources of Error / *9.26*
9.8 Target-caused Errors (Target Noise) / *9.26*
9.9 Other External Causes of Error / *9.37*
9.10 Internal Sources of Error / *9.42*
9.11 Summary of Sources of Error / *9.43*
9.12 Error Reduction Techniques / *9.46*

Chapter 10 The Radar Transmitter **10.1**
Thomas A. Weil and Merrill Skolnik
10.1 Introduction / *10.1*
10.2 Linear-beam Amplifiers / *10.4*
10.3 Magnetron / *10.14*
10.4 Crossed-field Amplifiers / *10.16*
10.5 Gyrotrons / *10.17*
10.6 Transmitter Spectrum Control / *10.19*
10.7 Grid-controlled Tubes / *10.21*

10.8 Modulators / *10.23*
10.9 Which RF Power Source to Use? / *10.25*

Chapter 11 Solid id-State Transmitters **11.1**
Michael T. Borkowski
11.1 Introduction / *11.1*
11.2 Advantages of Solid State / *11.1*
11.3 Solid-state Devices / *11.5*
11.4 Designing for the Solid-state Bottle Transmitter / *11.17*
11.5 Designing for the Solid-state Phased Array Transmitter / *11.24*
11.6 Solid-state System Examples / *11.37*

Chapter 12 Reflector Antennas **12.1**
Michael E. Cooley and Daniel Davis
12.1 Introduction / *12.7*
12.2 Basic Principles and Parameters / *12.3*
12.3 Reflector Antenna Architectures / *12.16*
12.4 Reflector Feeds / *12.25*
12.5 Reflector Antenna Analysis / *12.37*
12.6 Mechanical Design Considerations / *12.35*
Acknowledgments / *12.47*

Chapter 13 Phased Array Radar Antennas **13.1**
Joe Frank and John D. Richards
13.1 Introduction / *13.7*
13.2 Array Theory / *13.9*
13.3 Planar Arrays and Beam Steering / *13.15*
13.4 Aperture Matching and Mutual Coupling / *13.20*
13.5 Low-sidelobe Phased Arrays / *13.28*
13.6 Quantization Effects / *13.34*
13.7 Bandwidth of Phased Arrays / *13.38*
13.8 Feed Networks (Beamformers) / *13.46*
13.9 Phase Shifters / *13.57*
13.10 Solid-state Modules / *13.53*
13.11 Multiple Simultaneous Receive Beams / *13.54*
13.12 Digital Beamforming / *13.56*
13.13 Radiation Pattern Nulling / *13.57*
13.14 Calibration of Active Phased Array Antennas / *13.60*
13.15 Phased Array Systems / *13.62*

Chapter 14 Radar Cross Section **14.1**
Eugene F. Knott
14.1 Introduction / *14.1*

14.2 The Concept of Echo Power / *14.4*
14.3 RCS Prediction Techniques / *14.16*
14.4 RCS Measurement Techniques / *14.27*
14.5 Radar Echo Suppression / *14.36*

Chapter 15 Sea Clutter **15.1**
Lewis B. Wetzel
15.1 Introduction / *15.1*
15.2 The Sea Surface / *15.3*
15.3 Empirical Behavior of Sea Clutter / *15.7*
15.4 Theories and Models of Sea Clutter / *15.27*
15.5 Summary and Conclusions / *15.37*

Chapter 16 Ground Echo **16.1**
Richard K. Moore
16.1 Introduction / *16.1*
16.2 Parameters Affecting Ground Return / *16.4*
16.3 Theoretical Models and Their Limitations / *16.7*
16.4 Fading of Ground Echoes / *16.12*
16.5 Measurement Techniques for Ground Return / *16.19*
16.6 General Models for Scattering Coefficient (Clutter Models) / *16.29*
16.7 Scattering Coefficient Data / *16.35*
16.8 Polarimetry / *16.46*
16.9 Scattering Coefficient Data Near Grazing / *16.52*
16.10 Imaging Radar Interpretation / *16.55*

Chapter 17 Synthetic Aperture Radar **17.1**
Roger Sullivan
17.1 Basic Principle of SAR / *17.1*
17.2 Early History of SAR / *17.2*
17.3 Types of SAR / *17.2*
17.4 SAR Resolution / *17.6*
17.5 Key Aspects of SAR / *17.10*
17.6 SAR Image Quality / *17.16*
17.7 Summary of Key SAR Equations / *17.21*
17.8 Special SAR Applications / *17.22*

Chapter 18 Space-Based Remote Sensing Radars **18.1**
R. Keith Raney
18.1 Perspective / *18.1*
18.2 Synthetic Aperture Radar (SAR) / *18.5*
18.3 Altimeters / *18.29*
18.4 Planetary Radars / *18.43*
18.5 Scatterometers / *18.53*
18.6 Radar Sounders / *18.59*

Chapter 19 Meteorological Radar **19.1**
R. Jeffrey Keeler and Robert J. Serafin
19.1 Introduction / *19.1*
19.2 The Radar Equation for Meteorological Targets / *19.3*
19.3 Design Considerations / *19.6*
19.4 Signal Processing / *19.19*
19.5 Operational Applications / *19.25*
19.6 Research Applications / *19.33*

Chapter 20 HF Over-the-Horizon Radar **20.1**
James M. Headrick and Stuart J. Anderson
20.1 Introduction / *20.1*
20.2 The Radar Equation / *20.5*
20.3 Factors Influencing Skywave Radar Design / *20.7*
20.4 The Ionosphere and Radiowave Propagation / *20.13*
20.5 Waveforms for HF Radar / *20.21*
20.6 The Transmitting System / *20.23*
20.7 Radar Cross Section / *20.26*
20.8 Clutter: Echoes from the Environment / *20.29*
20.9 Noise, Interference, and Spectrum Occupancy / *20.40*
20.10 The Receiving System / *20.45*
20.11 Signal Processing and Tracking / *20.49*
20.12 Radar Resource Management / *20.54*
20.13 Radar Performance Modeling / *20.55*
Appendix: HF Surface Wave Radar / *20.70*

Chapter 21 Ground Penetrating Radar **21.1**
David Daniels
21.1 Introduction / *21.1*
21.2 Physics of Propagation in Materials / *21.6*
21.3 Modeling / *21.13*
21.4 Properties of Materials / *21.18*
21.5 GPR Systems / *21.20*
21.6 Modulation Techniques / *21.21*
21.7 Antennas / *21.24*
21.8 Signal and Image Processing / *21.30*
21.9 Applications / *21.35*
21.10 Licensing / *21.39*

Chapter 22 Civil Marine Radar **22.1**
Andy Norris
22.1 Introduction / *22.1*
22.2 The Challenges / *22.3*
22.3 International Standards / *22.7*
22.4 Technology / *22.10*
22.5 Target Tracking / *22.17*

22.6 User Interface / *22.19*
22.7 Integration with AIS / *22.23*
22.8 Radar Beacons / *22.25*
22.9 Validation Testing / *22.28*
22.10 Vessel Tracking Services / *22.29*
Appendix The Early Days of CMR / *22.31*
List of Maritime Radar-related Abbreviations / *22.33*
Acknowledgments / *22.34*

Chapter 23 Bistatic Radar **23.1**
Nicholas J. Willis
23.1 Concept and Definitions / *23.1*
23.2 Coordinate Systems / *23.3*
23.3 Bistatic Radar Equation / *23.4*
23.4 Applications / *23.9*
23.5 Bistatic Doppler / *23.14*
23.6 Target Location / *23.17*
23.7 Target Cross Section / *23.19*
23.8 Surface Clutter / *23.22*
23.9 Unique Problems and Requirements / *23.26*

Chapter 24 Electronic Counter-Countermeasures **24.1**
Alfonso Farina
24.1 Introduction / *24.1*
24.2 Terminology / *24.2*
24.3 Electronic Warfare Support Measures / *24.2*
24.4 Electronic Countermeasures / *24.5*
24.5 Objectives and Taxonomy of ECCM Techniques / *24.8*
24.6 Antenna-related ECCM / *24.10*
24.7 Transmitter-related ECCM / *24.31*
24.8 Receiver-related ECCM / *24.32*
24.9 Signal-processing-related ECCM / *24.33*
24.10 Operational-deployment Techniques / *24.36*
24.11 Application of ECCM Techniques / *24.37*
24.12 ECCM and ECM Efficacy / *24.54*
Acronym List / *24.56*
Acknowledgments / *24.58*

Chapter 25 Radar Digital Signal Processing **25.1**
James J. Alter and Jeffrey O. Coleman
25.1 Introduction / *25.1*
25.2 Receive Channel Processing / *25.2*
25.3 Transmit Channel Processing / *25.20*
25.4 DSP Tools / *25.22*
25.5 Design Considerations / *25.34*
25.6 Summary / *25.37*
Acknowledgments / *25.38*

Chapter 26 The Propagation Factor, F_p, in the Radar Equation **26.1**
Wayne L. Patterson
26.1 Introduction / *26.1*
26.2 The Earth's Atmosphere / *26.2*
26.3 Refraction / *26.3*
26.4 Standard Propagation / *26.4*
26.5 Anomalous Propagation / *26.6*
26.6 Propagation Modeling / *26.13*
26.7 EM System Assessment Programs / *26.18*
26.8 AREPS Radar System Assessment Model / *26.23*
26.9 AREPS Radar Displays / *26.25*

Index **1.1**

CONTRIBUTORS

James J. Alter *Naval Research Laboratory (CHAPTER 25)*
Stuart J. Anderson *Australian Defense Science and Technology Organisation (CHAPTER 20)*
W. G. Bath *The Johns Hopkins University Applied Physics Laboratory (CHAPTER 7)*
Michael T. Borkowski *Raytheon Company (CHAPTER 11)*
Jeffrey O. Coleman *Naval Research Laboratory (CHAPTER 25)*
Michael E. Cooley *Northrop Grumman, Electronic Systems (CHAPTER 12)*
David Daniels *ERA Technology (CHAPTER 21)*
Daniel Davis *Northrop Grumman Corporation (CHAPTER 12)*
James K. Day *Lockheed Martin Corporation (CHAPTER 3)*
Michael R. Ducoff *Lockheed Martin Corporation (CHAPTER 8)*
Alfonso Farina *SELEX Sistemi Integrati (CHAPTER 24)*
William G. Fedarko *Northrop Grumman Corporation (CHAPTER 4)*
Joe Frank *The Johns Hopkins University Applied Physics Laboratory (CHAPTER 13)*
Vilhelm Gregers-Hansen *Naval Research Laboratory (CHAPTER 2)*
James M. Headrick *Naval Research Laboratory, retired (CHAPTER 20)*
Dean D. Howard *Consultant to ITT Industries, Inc. (CHAPTER 9)*
R. Jeffrey Keeler *National Center for Atmospheric Research (CHAPTER 19)*
Eugene F. Knott *Tomorrow's Research (CHAPTER 14)*
Carlo Kopp *Monash University (CHAPTER 5)*
David Lynch, Jr. *DL Sciences, Inc. (CHAPTER 5)*
Richard K. Moore *The University of Kansas (CHAPTER 16)*
Andy Norris *Consultant in Navigation Systems (CHAPTER 22)*
Wayne L. Patterson *Space and Naval Warfare Systems Center (CHAPTER 26)*
Keith Raney *The Johns Hopkins University Applied Physics Laboratory (CHAPTER 18)*
John D. Richards *The Johns Hopkins University Applied Physics Laboratory (CHAPTER 13)*
Robert J. Serafin *National Center for Atmospheric Research (CHAPTER 19)*
William W. Shrader *Shrader Associates (CHAPTER 2)*
Merrill Skolnik *(CHAPTERS 1 and 10)*
Fred M. Staudaher *Naval Research Laboratory, retired (CHAPTER 3)*

John P. Stralka *Northrop Grumman Corporation (CHAPTER 4)*
Roger Sullivan *Institute for Defense Analyses (CHAPTER 17)*
Byron W. Tietjen *Lockheed Martin Corporation (CHAPTER 8)*
G. V. Trunk *The Johns Hopkins University Applied Physics Laboratory (CHAPTER 7)*
Thomas A. Weil *(CHAPTER 10)*
Lewis B. Wetzel *Naval Research Laboratory, retired (CHAPTER 15)*
Nicholas J. Willis *Technology Service Corporation, retired (CHAPTER 23)*
Michael E. Yeomans *Raytheon Company (CHAPTER 6)*

PREFACE

Radar is an important example of an electrical engineering system. In university engineering courses, the emphasis usually is on the basic tools of the electrical engineer such as circuit design, signals, solid state, digital processing, electronic devices, electromagnetics, automatic control, microwaves, and so forth. But in the real world of electrical engineering practice, these are only the techniques, piece parts, or subsystems that make up some type of system employed for a useful purpose. In addition to radar and other sensor systems, electrical engineering systems include communications, control, energy, information, industrial, military, navigation, entertainment, medical, and others. These are what the practice of electrical engineering is all about. Without them there would be little need for electrical engineers. However, the practicing engineer who is involved in producing a new type of electrical engineering system often has to depend on acquiring knowledge that was not usually covered in his or her engineering courses. The radar engineer, for example, has to understand the major components and subsystems that make up a radar, as well as how they fit together. The *Radar Handbook* attempts to help in this task. In addition to the radar system designer, it is hoped that those who are responsible for procuring new radar systems, those who utilize radars, those who maintain radar systems, and those who manage the engineers who do the above, also will find the *Radar Handbook* to be of help in fulfilling such tasks.

The third edition of the *Radar Handbook* is evidence that the development and application of radar for both civilian and military purposes continue to grow in both utility and in improved technology. Some of the many advances in radar since the previous edition include the following:

- The extensive use of digital methods for improved signal processing, data processing, decision making, flexible radar control, and multifunction radar
- Doppler weather radar
- Ground moving target indication, or GMTI
- An extensive experimental database describing low-angle land clutter, as obtained by MIT Lincoln Laboratory, that replaced the previously widely used clutter model that dated back to World War II
- The realization that microwave sea echo at low grazing angles is due chiefly to what are called “sea spikes”
- The active-aperture phased array radar system using solid-state modules, also called active electronically scanned arrays (AESA), which is attractive for some multifunction radar applications that need to manage both power and spatial coverage
- Planetary exploration with radar
- Computer-based methods for predicting radar propagation performance in realistic environments

- Operational use of HF over-the-horizon radar
- Improved methods for detecting moving targets in clutter, including space-time adaptive processing
- Operational use of inverse synthetic aperture radar for target recognition
- Interferometric synthetic aperture radar, or InSAR, to obtain the height of a resolved scatterer or to detect moving ground targets as well as provide a SAR image of a scene
- High precision space-based altimeters, with accuracy of a few centimeters, to measure the Earth's geoid
- Ultrawideband radar for ground penetrating and similar applications
- Improved high power, wide bandwidth klystron power sources based on clustered cavity resonators, as well as the multiple-beam klystron
- The appearance of wide bandgap semiconductors that allow better performance because of high power and high operating temperatures
- The availability of high-power millimeter-wave generators based on the gyroklystron
- Nonlinear FM pulse compression with low sidelobe levels
- The replacement, by the computer, of the operator as information extractor and decision maker

The above are not listed in any particular order, nor should they be considered a complete enumeration of radar developments since the appearance of the previous edition. There were also some radar topics in previous editions of the *Radar Handbook* that are of lesser interest and so were not included in this edition.

The chapter authors, who are experts in their particular field, were told to consider the reader of their chapter as being knowledgeable in the general subject of radar and even an expert in some other particular area of radar, but not necessarily knowledgeable about the subject of the particular chapter the author was writing.

It should be expected that with a book in print as long as the *Radar Handbook* has been, not all chapter authors from the previous editions would be available to do the third edition. Many of the previous authors have retired or are no longer with us. Sixteen of the twenty-six chapters in this edition have authors or coauthors who were not involved in the previous editions.

The hard work of preparing these chapters was done by the individual expert authors of the various chapters. Thus the value of the *Radar Handbook* is the result of the diligence and expertise of the authors who contributed their time, knowledge, and experience to make this handbook a useful addition to the desk of radar system engineers and all those people vital to the development, production, and employment of radar systems. I am deeply grateful to all the contributing authors for their fine work and the long hours they had to apply to their task. It is the chapter authors who make any handbook a success. My sincere thanks to them all.

As stated in the Preface of the previous edition, readers who wish to reference or quote material from the *Radar Handbook* are asked to mention the names of the individual chapter authors who produced the material.

MERRILL SKOLNIK
Baltimore, Maryland

Chapter 1

An Introduction and Overview of Radar*

Merrill Skolnik

1.1 RADAR IN BRIEF

Radar is an electromagnetic sensor for the detection and location of reflecting objects. Its operation can be summarized as follows:

- The radar radiates electromagnetic energy from an antenna to propagate in space.
- Some of the radiated energy is intercepted by a reflecting object, usually called a *target*, located at a distance from the radar.
- The energy intercepted by the target is reradiated in many directions.
- Some of the reradiated (echo) energy is returned to and received by the radar antenna.
- After amplification by a receiver and with the aid of proper signal processing, a decision is made at the output of the receiver as to whether or not a target echo signal is present. At that time, the target location and possibly other information about the target is acquired.

A common waveform radiated by a radar is a series of relatively narrow, rectangular-like pulses. An example of a waveform for a medium-range radar that detects aircraft might be described as a short pulse one millionth of a second in duration (one microsecond); the time between pulses might be one millisecond (so that the pulse repetition frequency is one kilohertz); the peak power from the radar transmitter might be one million watts (one megawatt); and with these numbers, the average power from the transmitter is one kilowatt. An average power of one kilowatt might be less than the power of the electric lighting usually found in a "typical" classroom. We assume this example radar might operate in the middle of the microwave† frequency range such as from 2.7 to 2.9 GHz, which is a typical frequency band for civil

* This chapter is a brief overview of radar for those not too familiar with the subject. For those who are familiar with radar, it can be considered a refresher.

† Microwaves are loosely defined as those frequencies where waveguides are used for transmission lines and where cavities or distributed circuits are used for resonant circuits rather than lumped-constant components. Microwave radars might be from about 400 MHz to about 40 GHz, but these limits are not rigid.

airport-surveillance radars. Its wavelength might be about 10 cm (rounding off, for simplicity). With the proper antenna such a radar might detect aircraft out to ranges‡ of 50 to 60 nmi, more or less. The echo power received by a radar from a target can vary over a wide range of values, but we arbitrarily assume a "typical" echo signal for illustrative purposes might have a power of perhaps 10^{-13} watts. If the radiated power is 10^6 watts (one megawatt), the ratio of echo signal power from a target to the radar transmitter power in this example is 10^{-19}, or the received echo is 190 dB less than the transmitted signal. That is quite a difference between the magnitude of the transmitted signal and a detectable received echo signal.

Some radars have to detect targets at ranges as short as the distance from behind home plate to the pitcher's mound in a baseball park (to measure the speed of a pitched ball), while other radars have to operate over distances as great as the distances to the nearest planets. Thus, a radar might be small enough to hold in the palm of one hand or large enough to occupy the space of many football fields.

Radar targets might be aircraft, ships, or missiles; but radar targets can also be people, birds, insects, precipitation, clear air turbulence, ionized media, land features (vegetation, mountains, roads, rivers, airfields, buildings, fences, and power-line poles), sea, ice, icebergs, buoys, underground features, meteors, aurora, spacecraft, and planets. In addition to measuring the range to a target as well as its angular direction, a radar can also find the relative velocity of a target either by determining the rate of change of the range measurement with time or by extracting the radial velocity from the doppler frequency shift of the echo signal. If the location of a moving target is measured over a period of time, the track, or trajectory, of the target can be found from which the absolute velocity of the target and its direction of travel can be determined and a prediction can be made as to its future location. Properly designed radars can determine the size and shape of a target and might even be able to recognize one type or class of target from another.

Basic Parts of a Radar. Figure 1.1 is a very elementary basic block diagram showing the subsystems usually found in a radar. The *transmitter*, which is shown here as a power amplifier, generates a suitable waveform for the particular job the radar is to perform. It might have an average power as small as milliwatts or as large as megawatts. (The average power is a far better indication of the capability of a radar's performance than is its peak power.) Most radars use a short pulse waveform so that a single antenna can be used on a time-shared basis for both transmitting and receiving.

The function of the *duplexer* is to allow a single antenna to be used by protecting the sensitive receiver from burning out while the transmitter is on and by directing the received echo signal to the receiver rather than to the transmitter.

The *antenna* is the device that allows the transmitted energy to be propagated into space and then collects the echo energy on receive. It is almost always a directive antenna, one that directs the radiated energy into a narrow beam to concentrate the power as well as to allow the determination of the direction to the target. An antenna that produces a narrow directive beam on transmit usually has a large area on receive to allow the collection of weak echo signals from the target. The antenna not only concentrates the energy on transmit and collects the echo energy on receive, but it also acts as a spatial filter to provide angle resolution and other capabilities.

‡ In radar, *range* is the term generally used to mean distance from the radar to the target. Range is also used here in some of its other dictionary definitions.

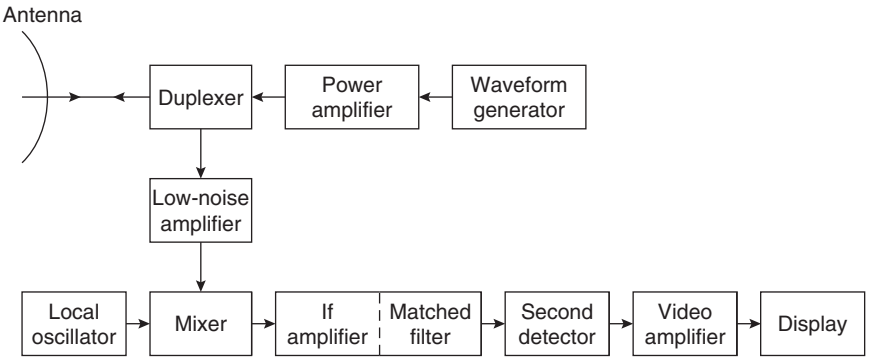


FIGURE 1.1 Block diagram of a simple radar employing a power amplifier as the transmitter in the upper portion of the figure and a superheterodyne receiver in the lower portion of the figure

The *receiver* amplifies the weak received signal to a level where its presence can be detected. Because noise is the ultimate limitation on the ability of a radar to make a reliable detection decision and extract information about the target, care is taken to insure that the receiver produces very little noise of its own. At the microwave frequencies, where most radars are found, the noise that affects radar performance is usually from the first stage of the receiver, shown here in Figure 1.1 as a *low-noise amplifier*. For many radar applications where the limitation to detection is the unwanted radar echoes from the environment (called *clutter*), the receiver needs to have a large enough dynamic range so as to avoid having the clutter echoes adversely affect detection of wanted moving targets by causing the receiver to saturate. The *dynamic range* of a receiver, usually expressed in decibels, is defined[1] as the ratio of the maximum to the minimum signal input power levels over which the receiver can operate with some specified performance. The maximum signal level might be set by the nonlinear effects of the receiver response that can be tolerated (for example, the signal power at which the receiver begins to saturate), and the minimum signal might be the minimum detectable signal. The *signal processor*, which is often in the IF portion of the receiver, might be described as being the part of the receiver that separates the desired signal from the undesired signals that can degrade the detection process. Signal processing includes the *matched filter* that maximizes the output signal-to-noise ratio. Signal processing also includes the doppler processing that maximizes the signal-to-clutter ratio of a moving target when clutter is larger than receiver noise, and it separates one moving target from other moving targets or from clutter echoes. The *detection decision* is made at the output of the receiver, so a target is declared to be present when the receiver output exceeds a predetermined threshold. If the threshold is set too low, the receiver noise can cause excessive false alarms. If the threshold is set too high, detections of some targets might be missed that would otherwise have been detected. The criterion for determining the level of the decision threshold is to set the threshold so it produces an acceptable predetermined average rate of false alarms due to receiver noise.

After the detection decision is made, the track of a target can be determined, where a track is the locus of target locations measured over time. This is an example of *data processing*. The processed target detection information or its track might be displayed to an operator; or the detection information might be used to automatically guide a

missile to a target; or the radar output might be further processed to provide other information about the nature of the target. The *radar control* insures that the various parts of a radar operate in a coordinated and cooperative manner, as, for example, providing timing signals to various parts of the radar as required.

The radar engineer has as resources *time* that allows good doppler processing, *bandwidth* for good range resolution, *space* that allows a large antenna, and *energy* for long range performance and accurate measurements. External factors affecting radar performance include the *target characteristics; external noise* that might enter via the antenna; unwanted *clutter* echoes from land, sea, birds, or rain; *interference* from other electromagnetic radiators; and propagation effects due to the earth's surface and atmosphere. These factors are mentioned to emphasize that they can be highly important in the design and application of a radar.

Radar Transmitters. The radar transmitter must not only be able to generate the peak and average powers required to detect the desired targets at the maximum range, but also has to generate a signal with the proper waveform and the stability needed for the particular application. Transmitters may be oscillators or amplifiers, but the latter usually offer more advantages.

There have been many types of radar power sources used in radar (Chapter 10). The magnetron power oscillator was at one time very popular, but it is seldom used except for civil marine radar (Chapter 22). Because of the magnetron's relatively low average power (one or two kilowatts) and poor stability, other power sources are usually more appropriate for applications requiring long-range detection of small moving targets in the presence of large clutter echoes. The magnetron power oscillator is an example of what is called a *crossed-field tube*. There is also a related *crossed-field amplifier* (CFA) that has been used in some radars in the past, but it also suffers limitations for important radar applications, especially for those requiring detection of moving targets in clutter. The high-power klystron and the traveling wave tube (TWT) are examples of what are called *linear beam tubes*. At the high powers often employed by radars, both tubes have suitably wide bandwidths as well as good stability as needed for doppler processing, and both have been popular.

The *solid-state amplifier*, such as the transistor, has also been used in radar, especially in phased arrays. Although an individual transistor has relatively low power, each of the many radiating elements of an array antenna can utilize multiple transistors to achieve the high power needed for many radar applications. When solid-state transistor amplifiers are used, the radar designer has to be able to accommodate the high duty cycle at which these devices have to operate, the long pulses they must use that require pulse compression, and the multiple pulses of different widths to allow detection at short as well as long range. Thus the use of solid-state transmitters can have an effect on other parts of the radar system. At millimeter wavelengths very high power can be obtained with the *gyrotron*, either as an amplifier or as an oscillator. The *grid-control vacuum tube* was used to good advantage for a long time in UHF and lower frequency radars, but there has been less interest in the lower frequencies for radar.

Although not everyone might agree, some radar system engineers—if given a choice—would consider the klystron amplifier as the prime candidate for a high-power modern radar if the application were suitable for its use.

Radar Antennas. The antenna is what connects the radar to the outside world (Chapters 12 and 13). It performs several purposes: (1) concentrates the radiated energy

on transmit; that is, it is directive and has a narrow beamwidth; (2) collects the received echo energy from the target; (3) provides a measurement of the angular direction to the target; (4) provides spatial resolution to resolve (or separate) targets in angle; and (5) allows the desired volume of space to be observed. The antenna can be a mechanically scanned parabolic reflector, a mechanically scanned planar phased array, or a mechanically scanned end-fire antenna. It can be an electronically scanned phased array using a single transmitter with either a corporate feed or a space-feed configuration to distribute the power to each antenna element or an electronically scanned phased array employing at each antenna element a small solid-state "miniature" radar (also called an *active aperture phased array*). Each type of antenna has its particular advantages and limitations. Generally, the larger the antenna the better, but there can be practical constraints that limit its size.

1.2 TYPES OF RADARS

Although there is no single way to characterize a radar, here we do so by means of what might be the major feature that distinguishes one type of radar from another.

Pulse radar. This is a radar that radiates a repetitive series of almost-rectangular pulses. It might be called the canonical form of a radar, the one usually thought of as a radar when nothing else is said to define a radar.

High-resolution radar. High resolution can be obtained in the range, angle, or doppler velocity coordinates, but high resolution usually implies that the radar has high range resolution. Some high-resolution radars have range resolutions in terms of fractions of a meter, but it can be as small as a few centimeters.

Pulse compression radar. This is a radar that uses a long pulse with internal modulation (usually frequency or phase modulation) to obtain the energy of a long pulse with the resolution of a short pulse.

Continuous wave (CW) radar. This radar employs a continuous sine wave. It almost always uses the doppler frequency shift for detecting moving targets or for measuring the relative velocity of a target.

FM-CW radar. This CW radar uses frequency modulation of the waveform to allow a range measurement.

Surveillance radar. Although a dictionary might not define *surveillance* this way, a surveillance radar is one that detects the presence of a target (such as an aircraft or a ship) and determines its location in range and angle. It can also observe the target over a period of time so as to obtain its track.

Moving target indication (MTI). This is a pulse radar that detects moving targets in clutter by using a low pulse repetition frequency (PRF) that usually has no range ambiguities. It does have ambiguities in the doppler domain that result in so-called blind speeds.

Pulse doppler radar. There are two types of pulse doppler radars that employ either a high or medium PRF pulse radar. They both use the doppler frequency shift to extract moving targets in clutter. A *high PRF pulse doppler radar* has no ambiguities (blind speeds) in doppler, but it does have range ambiguities. A *medium PRF pulse doppler radar* has ambiguities in both range and doppler.

Tracking radar. This is a radar that provides the track, or trajectory, of a target. Tracking radars can be further delineated as STT, ADT, TWS, and phased array trackers as described below:

Single Target Tracker (STT). Tracks a single target at a data rate high enough to provide accurate tracking of a maneuvering target. A revisit time of 0.1 s (data rate of 10 measurements per second) might be "typical." It might employ the monopulse tracking method for accurate tracking information in the angle coordinate.

Automatic detection and tracking (ADT). This is tracking performed by a surveillance radar. It can have a very large number of targets in track by using the measurements of target locations obtained over multiple scans of the antenna. Its data rate is not as high as the STT. Revisit times might range from one to 12 seconds, depending on the application.

Track-while-scan (TWS). Usually a radar that provides surveillance over a narrow region of angle in one or two dimensions, so as to provide at a rapid update rate location information on all targets within a limited angular region of observation. It has been used in the past for ground-based radars that guide aircraft to a landing, in some types of weapon-control radars, and in some military airborne radars.

Phased array tracker. An electronically scanned phased array can (almost) "continuously" track more than one target at a high data rate. It can also simultaneously provide the lower data rate tracking of multiple targets similar to that performed by ADT.

Imaging radar. This radar produces a two-dimensional image of a target or a scene, such as a portion of the surface of the earth and what is on it. These radars usually are on moving platforms.

Sidelooking airborne radar (SLAR). This airborne sidelooking imaging radar provides high resolution in range and obtains suitable resolution in angle by using a narrow beamwidth antenna.

Synthetic aperture radar (SAR). SAR is a coherent* imaging radar on a moving vehicle that uses the phase information of the echo signal to obtain an image of a scene with high resolution in both range and cross-range. High range resolution is often obtained using pulse compression.

Inverse synthetic aperture radar (ISAR). ISAR is a coherent imaging radar that uses high resolution in range and the relative motion of the target to obtain high resolution in the doppler domain that allows resolution in the cross-range dimension to be obtained. It can be on a moving vehicle or it can be stationary.

Weapon control radar. This name is usually applied to a single-target tracker used for defending against air attack.

Guidance radar. This is usually a radar on a missile that allows the missile to "home in," or guide itself, to a target.

Weather (meteorological) observation. Such radars detect, recognize, and measure precipitation rate, wind speed and direction, and observe other weather effects

* *Coherent* implies that the phase of the radar signal is used as an important part of the radar process.

important for meteorological purposes. These may be special radars or another function of surveillance radars.

Doppler weather radar. This is a weather observation radar that employs the doppler frequency shift caused by moving weather effects to determine the wind; the wind shear (when the wind blows in different directions), which can indicate a dangerous weather condition such as a tornado or a downburst of wind; as well as other meteorological effects.

Target recognition. In some cases, it might be important to recognize the type of target being observed by radar (e.g., an automobile rather than a bird), or to recognize the particular type of target (an automobile rather than a truck, or a starling rather than a sparrow), or to recognize one class of target from another (a cruise ship rather than a tanker). When used for military purposes, it is usually called a *noncooperative target recognition* (NCTR) radar, as compared to a cooperative recognition system such as IFF (identification friend or foe), which is not a radar. When target recognition involves some part of the natural environment, the radar is usually known as a *remote sensing* (of the environment) radar.

Multifunction radar. If each of the above radars were thought of as providing some radar function, then a multifunction radar is one designed to perform more than one such function—usually performing one function at a time on a time-shared basis.

There are many other ways to describe radars, including land, sea, airborne, spaceborne, mobile, transportable, air-traffic control, military, ground-penetrating, ultrawideband, over the horizon, instrumentation, laser (or lidar), by the frequency band at which they operate (UHF, L, S, and so on), by their application, and so forth.

1.3 INFORMATION AVAILABLE FROM A RADAR

Detection of targets is of little value unless some information about the target is obtained as well. Likewise, target information without target detection is meaningless.

Range. Probably the most distinguishing feature of a conventional radar is its ability to determine the range to a target by measuring the time it takes for the radar signal to propagate at the speed of light out to the target and back to the radar. No other sensor can measure the distance to a remote target at long range with the accuracy of radar (basically limited at long ranges by the accuracy of the knowledge of the velocity of propagation). At modest ranges, the precision can be a few centimeters. To measure range, some sort of *timing mark* must be introduced on the transmitted waveform. A timing mark can be a short pulse (an amplitude modulation of the signal), but it can also be a distinctive modulation of the frequency or phase. The accuracy of a range measurement depends on the radar signal bandwidth: the wider the bandwidth, the greater the accuracy. Thus *bandwidth* is the basic measure of range accuracy.

Radial Velocity. The radial velocity of a target is obtained from the rate of change of range over a period of time. It can also be obtained from the measurement of the doppler frequency shift. An accurate measurement of radial velocity requires time. Hence *time* is the basic parameter describing the quality of a radial velocity measurement. The speed of a moving target and its direction of travel can be obtained from its track, which can be found from the radar measurements of the target location over a period of time.

Angular Direction. One method for determining the direction to a target is by determining the angle where the magnitude of the echo signal from a scanning antenna is maximum. This usually requires an antenna with a narrow beamwidth (a high-gain antenna). An air-surveillance radar with a rotating antenna beam determines angle in this manner. The angle to a target in one angular dimension can also be determined by using two antenna beams, slightly displaced in angle, and comparing the echo amplitude received in each beam. Four beams are needed to obtain the angle measurement in both azimuth and elevation. The monopulse tracking radar discussed in Chapter 9 is a good example. The accuracy of an angle measurement depends on the *electrical size of the antenna*; i.e., the size of the antenna given in wavelengths.

Size and Shape. If the radar has sufficient resolution capability in range or angle, it can provide a measurement of the target extent in the dimension of high resolution. Range is usually the coordinate where resolution is obtained. Resolution in cross range (given by the range multiplied by the antenna beamwidth) can be obtained with very narrow beamwidth antennas. However, the angular width of an antenna beam is limited, so the cross-range resolution obtained by this method is not as good as the range resolution. Very good resolution in the cross-range dimension can be obtained by employing the doppler frequency domain, based on SAR (synthetic aperture radar) or ISAR (inverse synthetic aperture radar systems), as discussed in Chapter 17. There needs to be relative motion between the target and the radar in order to obtain the cross-range resolution by SAR or ISAR. With sufficient resolution in both range and cross-range, not only can the size be obtained in two orthogonal coordinates, but the target shape can sometimes be discerned.

The Importance of Bandwidth in Radar. Bandwidth basically represents information; hence, it is very important in many radar applications. There are two types of bandwidth encountered in radar. One is the *signal bandwidth*, which is the bandwidth determined by the signal pulse width or by any internal modulation of the signal. The other is *tunable bandwidth*. Generally, the signal bandwidth of a simple pulse of sine wave of duration τ is $1/\tau$. (Pulse compression waveforms, discussed in Chapter 8, can have much greater bandwidth than the reciprocal of their pulse width.) Large bandwidth is needed for resolving targets in range, for accurate measurement of range to a target, and for providing a limited capability to recognize one type of target from another. High range resolution also can be useful for reducing the degrading effects of what is known as *glint* in a tracking radar, for measuring the altitude of an aircraft based on the difference in time delay (range) between the two-way direct signal from radar to target and the two-way surface-scattered signal from radar to surface to target (also called *multipath height finding*), and in increasing the target-signal-to-clutter ratio. In military systems, high range resolution may be employed for counting the number of aircraft flying in close formation and for recognizing and thwarting some types of deception countermeasures.

Tunable bandwidth offers the ability to change (tune) the radar signal frequency over a wide range of the available spectrum. This can be used for reducing mutual interference among radars that operate in the same frequency band, as well as in attempting to make hostile electronic countermeasures less effective. The higher the operating frequency the easier it is to obtain wide signal and wide tunable bandwidth.

A limitation on the availability of bandwidth in a radar is the control of the spectrum by government regulating agencies (in the United States, the Federal Communication

Commission, and internationally, the International Telecommunications Union). After the success of radar in World War II, radar was allowed to operate over about one-third of the microwave spectrum. This spectrum space has been reduced considerably over the years with the advent of many commercial users of the spectrum in the age of "wireless" and other services requiring the electromagnetic spectrum. Thus, the radar engineer is increasingly experiencing smaller available spectrum space and bandwidth allocations that can be vital for the success of many radar applications.

Signal-to-Noise Ratio. The accuracy of all radar measurements, as well as the reliable detection of targets depends on the ratio E/N_o, where E is the total energy of the received signal that is processed by the radar and N_o is the noise power per unit bandwidth of the receiver. Thus E/N_o is an important measure of the capability of a radar.

Operation with More Than One Frequency. There may be important benefits when a radar is able to operate at more than one frequency.[2] *Frequency agility* usually refers to the use of multiple frequencies on a pulse-to-pulse basis. *Frequency diversity* usually relates to the use of multiple frequencies that are widely separated, sometimes in more than one radar band. Frequency diversity might operate at each frequency simultaneously or almost simultaneously. It has been used in almost all civilian air-traffic control radars. Pulse-to-pulse frequency agility, however, is not compatible with the use of doppler processing (to detect moving targets in clutter), but frequency diversity can be compatible. The frequency range in both agility and in diversity operations is much greater than the inherent bandwidth of a pulse of width τ.

Elevation Null Filling. Operation of a radar at a single frequency can result in a lobed structure to the elevation pattern of an antenna due to the interference between the direct signal (radar to target) and the surface-scattered signal (radar to earth's surface to target). By a lobed structure, we mean that there will be reduced coverage at some elevation angles (nulls) and increased signal strength at other angles (lobes). A change in frequency will change the location of the nulls and lobes so that by operating over a wide frequency range, the nulls in elevation can be filled in, and the radar will be less likely to lose a target echo signal. For example, measurements with a wideband experimental radar known as Senrad, which could operate from 850 to 1400 MHz, showed that when only a single frequency was used, the blip-scan ratio (the experimentally measured single-scan probability of detection) was found to be 0.78 under a particular set of observations. When the radar operated at four different widely separated frequencies, the blip-scan ratio was 0.98—a highly significant increase due to frequency diversity.[2]

Increased Target Detectability. The radar cross section of a complex target such as an aircraft can vary greatly with a change in frequency. At some frequencies, the radar cross section will be small and at others it will be large. If a radar operates at a single frequency, it might result in a small target echo and, therefore, a missed detection. By operating at a number of different frequencies, the cross section will vary and can be small or large; but a successful detection becomes more likely than if only a single frequency were used. This is one reason that almost all air-traffic control radars operate with two frequencies spaced wide enough apart in frequency to insure that target echoes are decorrelated and, therefore, increase the likelihood of detection.

Reduced Effectiveness of Hostile Countermeasures. Any military radar that is successful can expect a hostile adversary to employ countermeasures to reduce its effectiveness. Operating over a wide range of frequencies makes countermeasures more difficult than if operation is at only one frequency. Against noise jamming, changing frequency in an unpredictable manner over a wide range of frequencies causes the jammer to have to spread its power over a wide frequency range and will, therefore, reduce the hostile jamming signal strength over the bandwidth of the radar signal. Frequency diversity over a wide band also makes it more difficult (but not impossible) for a hostile intercept receiver or an antiradiation missile to detect and locate a radar signal.

The Doppler Shift in Radar. The importance of the doppler frequency shift began to be appreciated for pulse radar shortly after World War II and became an increasingly important factor in many radar applications. Modern radar would be much less interesting or useful if the doppler effect didn't exist. The doppler frequency shift f_d can be written as

$$f_d = 2v_r/\lambda = (2v\cos\theta)/\lambda \tag{1.1}$$

where $v_r = v \cos \theta$ is the relative velocity of the target (relative to the radar) in m/s, v is the absolute velocity of the target in m/s, λ is the radar wavelength in m, and θ is the angle between the target's direction and the radar beam. To an accuracy of about 3 percent, the doppler frequency in hertz is approximately equal to v_r (kt) divided by λ (m).

The doppler frequency shift is widely used to separate moving targets from stationary clutter, as discussed in Chapters 2 through 5. Such radars are known as MTI (moving target indication), AMTI (airborne MTI), and pulse doppler. All modern air-traffic control radars, all important military ground-based and airborne air-surveillance radars, and all military airborne fighter radars take advantage of the doppler effect. Yet in WWII, none of these pulse radar applications used doppler. The CW (continuous wave) radar also employs the doppler effect for detecting moving targets, but CW radar for this purpose is not as popular as it once was. The HF OTH radar (Chapter 20) could not do its job of detecting moving targets in the presence of large clutter echoes from the earth's surface without the use of doppler.

Another significant application of radar that depends on the doppler shift is observation of the weather, as in the Nexrad radars of the U.S. National Weather Service (Chapter 19) mentioned earlier in this chapter.

Both the SAR and ISAR can be described in terms of their use of the doppler frequency shift (Chapter 17). The airborne doppler navigator radar is also based on the doppler shift. The use of doppler in a radar generally places greater demands on the stability of the radar transmitter, and it increases the complexity of the signal processing; yet these requirements are willingly accepted in order to achieve the significant benefits offered by doppler. It should also be mentioned that the doppler shift is the key capability of a radar that can measure speed, as by its diligent use by traffic police for maintaining vehicle speed limits and in other velocity measuring applications.

1.4 THE RADAR EQUATION

The radar range equation (or radar equation, for short) not only serves the very useful purpose of estimating the range of a radar as a function of the radar characteristics,

but also is quite useful as a guide for designing a radar system. The simple form of the radar equation may be written as

$$P_r = \frac{P_t G_t}{4\pi R^2} \times \frac{\sigma}{4\pi R^2} \times A_e \tag{1.2}$$

The right-hand side has been written as the product of three factors to represent the physical processes that take place. The first factor on the right is the *power density at a distance R* from a radar that radiates a power P_t from an antenna of gain G_t. The numerator, σ, of the second factor is the radar cross section of the target. It has the unit of area (for example, square meters) and is a measure of the energy redirected by the target back in the direction of the radar. The denominator of the second factor accounts for the divergence of the echo signal on its return path back to the radar. The product of the first two factors represents the power per unit area returned to the radar antenna. Note that the radar cross section of a target, σ, *is defined by this equation*. The receiving antenna of effective area A_e collects a portion P_r of the echo power returned to the radar. If the maximum radar range, $R_{\max}$, is defined as occurring when the received signal is equal to the minimum detectable signal of the radar, $S_{\min}$, the simple form of the radar equation becomes

$$R_{\max}^4 = \frac{P_t G_t A_e \sigma}{(4\pi)^2 S_{\min}} \tag{1.3}$$

Generally, most radars use the same antenna for both transmitting and receiving. From antenna theory, there is a relation between the gain G_t of the antenna on transmit and its effective area A_e on receive, which is $G_t = 4\pi A_e / \lambda^2$, where λ is the wavelength of the radar signal. Substituting this into Eq. 1.3 provides two other useful forms of the radar equation (not shown here): one that represents the antenna only by its gain and the other that represents the antenna only by its effective area.

The simple form of the radar equation is instructive, but not very useful since it leaves out many things. The minimum detectable signal, $S_{\min}$, is limited by receiver noise and can be expressed as

$$S_{\min} = kT_o B F_n (S/N)_1 \tag{1.4}$$

In this expression, kT_oB is the so-called thermal noise from an ideal ohmic conductor, where k = Boltzmann's constant, T_o is the standard temperature of 290 K, and B = receiver bandwidth (usually that of the IF stage of the superheterodyne receiver). The product kT_o is equal to 4×10^{-21} W/Hz. To account for the additional noise introduced by a practical (nonideal) receiver, the thermal noise expression is multiplied by the noise figure F_n of the receiver, defined as the noise out of a practical receiver to the noise out of an ideal receiver. For a received signal to be detectable, it has to be larger than the receiver noise by a factor denoted here as $(S/N)_1$. This value of signal-to-noise ratio $(S/N)_1$ is that required if only one pulse is present. It has to be large enough to obtain the required probability of false alarm (due to noise crossing the receiver threshold) and the required probability of detection (as can be found in various radar texts[3,4]). Radars, however, generally process more than one pulse before making a detection decision. We assume the radar waveform is a repetitive series of rectangular-like pulses. These pulses are integrated (added together) before a detection decision

is made. To account for these added signals, the numerator of the radar equation is multiplied by a factor $nE_i(n)$, where $E_i(n)$ is the efficiency in adding together n pulses. This value can also be found in standard texts.

The power P_t is the peak power of a radar pulse. The average power, P_{av}, is a better measure of the ability of a radar to detect targets, so it is sometimes inserted into the radar equation using $P_t = P_{\text{av}}/f_p\tau$, where f_p is the pulse repetition frequency of the pulse radar and τ is the pulse duration. The surface of the earth and the earth's atmosphere can drastically affect the propagation of electromagnetic waves and change the coverage and capabilities of a radar. In the radar equation, these propagation effects are accounted for by a factor F^4 in the numerator of the radar equation, as discussed in Chapter 26. With the above substituted into the simple form of the radar equation we get

$$R_{\max}^4 = \frac{P_{\text{av}} G A_e \sigma n E_i(n) F^4}{(4\pi)^2 k T_o F_n f_p (S/N)_1 L_s} \tag{1.5}$$

In the above equation, it was assumed in its derivation that $B\tau \approx 1$, which is generally applicable in radar. A factor L_s (greater than unity), called the system losses, has been inserted to account for the many ways that loss can occur in a radar. This loss factor can be quite large. If the system loss is ignored, it might result in a very large error in the estimated range predicted by the radar equation. (A loss of from 10 dB to may be 20 dB is not unusual when *all* radar system loss factors are taken into account.)

Equation 1.5 applies for a radar that observes a target long enough to receive n pulses. More fundamentally, it applies for a radar where the time on target t_o is equal to n/f_p. An example is a tracking radar that continuously observes a single target for a time t_o. This equation, however, needs to be modified for a surveillance radar that observes an angular volume Ω with a revisit time t_s. (Air traffic control radars might have a revisit time of from 4 to 12 s.) Thus, a surveillance radar has the additional constraint that it must cover an angular volume Ω in a given time t_s. The revisit time t_s is equal to $t_o(\Omega/\Omega_o)$, where $t_o = n/f_p$ and Ω_o, the solid beamwidth of the antenna (steradians), is approximately related to the antenna gain G by $G = 4\pi/\Omega_o$. Therefore, when n/f_p in Eq. 1.5 is replaced with its equal $4\pi t_s/\text{G}\Omega$, the radar equation for a surveillance radar is obtained as

$$R_{\max}^4 = \frac{P_{\text{av}} A_e \sigma E_i(n) F^4}{4\pi k T_o F_n (S/N)_1 L_s} \times \frac{t_s}{\Omega} \tag{1.6}$$

The radar designer has little control over the revisit time t_s or the angular coverage Ω, which are determined mainly by the job the radar has to perform. The radar cross section also is determined by the radar application. If a large range is required of a surveillance radar, the radar must have the necessary value of the product $P_{\text{av}}A_e$. For this reason, a common measure of the capability of a surveillance radar is its *power-aperture product*. Note that the radar frequency does not appear *explicitly* in the surveillance radar equation. The choice of frequency, however, will enter *implicitly* in other ways.

Just as the radar equation for a surveillance radar is different from the conventional radar equation of Eq. 1.5 or the simple radar equation of Eq. 1.2, each particular application of a radar generally has to employ a radar equation tailored to that specific application. When the radar echoes from land, sea, or weather clutter are greater than the receiver noise, the radar equation has to be modified to account for clutter being the limitation to

detection rather than receiver noise. It can happen that the detection capability of a radar might be limited by clutter in some regions of its coverage and be limited by receiver noise in other regions. This can result in two different sets of radar characteristics, one optimized for noise and the other optimized for clutter; and compromises usually have to be made in radar design. A different type of radar equation is also required when the radar capability is limited by hostile noise jamming.

1.5 RADAR FREQUENCY LETTER-BAND NOMENCLATURE

It is not always convenient to use the exact numerical frequency range over which a particular type of radar operates. With many military radars, the exact operating frequency range of a radar is usually not disclosed. Thus, the use of letters to designate radar operating bands has been very helpful. The IEEE (Institute of Electrical and Electronic Engineers) has officially standardized the radar letter-band nomenclature, as summarized in Table 1.1.

Comments on the table.[6] The International Telecommunications Union (ITU) assigns specific portions of the electromagnetic spectrum for radiolocation (radar) use as shown in the third column, which applies to ITU Region 2 that includes North and South America. Slight differences occur in the other two ITU Regions. Thus an L-band radar can only operate within the frequency range from 1215 MHz to 1400 MHz, and even within this range, there may be restrictions. Some of the indicated ITU bands are restricted in their usage; for example, the band between 4.2 and 4.4 GHz is reserved

TABLE 1.1 IEEE Standard Letter Designations for Radar-Frequency Bands[5]

Band Designation	Nominal Frequency Range	Specific Frequency Ranges for Radar Based on ITU Frequency Assignments for Region 2
HF	3 MHz–30 MHz	
VHF	30–300 MHz	138–144 MHz
		216–225 MHz
UHF	300–1000 MHz	420–450 MHz
		890–942 MHz
L	1.0–2.0 GHz	1215–1400 MHz
S	2.0–4.0 GHz	2.3–2.5 GHz
		2.7–3.7 GHz
C	4.0–8.0 GHz	4.2–4.4 GHz
		5.25–5.925 GHz
X	8.0–12.0 GHz	8.5–10.68 GHz
K_u	12.0–18.0 GHz	13.4–14.0 GHz
		15.7–17.7 GHz
K	18.0–27.0 GHz	24.05–24.25 GHz
		24.65–24.75 GHz
K_a	27.0–40.0 GHz	33.4–36.0 GHz
V	40.0–75 GHz	59.0–64.0 GHz
W	75.0–110 GHz	76.0–81 GHz
		92.0–100 GHz

(with few exceptions) for airborne radar altimeters. There are no official ITU allocations for radar in the HF band, but most HF radars share frequencies with other electromagnetic services. The letter-band designation for millimeter wave radars is *mm*, and there are several frequency bands allocated to radar in this region, but they have not been listed here. Although the official ITU description of millimeter waves is from 30 to 300 GHz, in reality, the technology of radars at K_a band* is much closer to the technology of microwave frequencies than to the technology of W band. The millimeter wave radar frequencies are often considered by those who work in this field to have a lower bound of 40 GHz rather than the "legal" lower bound of 30 GHz in recognition of the significant difference in technology and applications that is characteristic of millimeter wave radar. *Microwaves* have not been defined in this standard, but this term generally applies to radars that operate from UHF to K_a band. The reason that these letter designations might not be easy for the non-radar engineer to recognize is that they were originally selected to describe the radar bands used in World War II. Secrecy was important at that time so the letters selected to designate the various bands made it hard to guess the frequencies to which they apply. Those who work around radar, however, seldom have a problem with the usage of the radar letter bands.

Other letter bands have been used for describing the electromagnetic spectrum; but they are not suitable for radar and should never be used for radar. One such designation uses the letters A, B, C, etc., originally devised for conducting electronic countermeasure exercises.[7] The IEEE Standard mentioned previously states that these "are not consistent with radar practice and shall not be used to describe radar-frequency bands." Thus, there may be D-band jammers, but never D-band radars.

1.6 EFFECT OF OPERATING FREQUENCY ON RADAR

Radars have been operated at frequencies as low as 2 MHz (just above the AM broadcast band) and as high as several hundred GHz (millimeter wave region). More usually, radar frequencies might be from about 5 MHz to over 95 GHz. This is a very large extent of frequencies, so it should be expected that radar technology, capabilities, and applications will vary considerably depending on the frequency range at which a radar operates. Radars at a particular frequency band usually have different capabilities and characteristics than radars in other frequency bands. Generally, long range is easier to achieve at the lower frequencies because it is easier to obtain high-power transmitters and physically large antennas at the lower frequencies. On the other hand, at the higher radar frequencies, it is easier to achieve accurate measurements of range and location because the higher frequencies provide wider bandwidth (which determines range accuracy and range resolution) as well as narrower beam antennas for a given physical size antenna (which determines angle accuracy and angle resolution). In the following, the applications usually found in the various radar bands are briefly indicated. The differences between adjacent bands, however, are seldom sharp in practice, and overlap in characteristics between adjacent bands is likely.

* The wavelengths of K_a band range from 8.3 mm to 9 mm, which qualifies them under the "legal" definition of millimeters, but just barely.

HF (3 to 30 MHz). The major use of the HF band for radar (Chapter 20) is to detect targets at long ranges (nominally out to 2000 nmi) by taking advantage of the refraction of HF energy by the ionosphere that lies high above the surface of the earth. Radio amateurs refer to this as *short-wave propagation* and use it to communicate over long distances. The targets for such HF radars might be aircraft, ships, and ballistic missiles, as well as the echo from the sea surface itself that provides information about the direction and speed of the winds that drive the sea.

VHF (30 to 300 MHz). At the beginning of radar development in the 1930s, radars were in this frequency band because these frequencies represented the frontier of radio technology at that time. It is a good frequency for long range air surveillance or detection of ballistic missiles. At these frequencies, the reflection coefficient on scattering from the earth's surface can be very large, especially over water, so the constructive interference between the direct signal and the surface-reflected signal can increase significantly the range of a VHF radar. Sometimes this effect can almost double the radar's range. However, when there is constructive interference that increases the range, there can be destructive interference that decreases the range due to the deep nulls in the antenna pattern in the elevation plane. Likewise, the destructive interference can result in poor low-altitude coverage. Detection of moving targets in clutter is often better at the lower frequencies when the radar takes advantage of the doppler frequency shift because doppler ambiguities (that cause blind speeds) are far fewer at low frequencies. VHF radars are not bothered by echoes from rain, but they can be affected by multiple-time-around echoes from meteor ionization and aurora. The radar cross section of aircraft at VHF is generally larger than the radar cross section at higher frequencies. VHF radars frequently cost less compared to radars with the same range performance that operate at higher frequencies.

Although there are many attractive advantages of VHF radars for long-range surveillance, they also have some serious limitations. Deep nulls in elevation and poor low-altitude coverage have been mentioned. The available spectral widths assigned to radar at VHF are small so range resolution is often poor. The antenna beamwidths are usually wider than at microwave frequencies, so there is poor resolution and accuracy in angle. The VHF band is crowded with important civilian services such as TV and FM broadcast, further reducing the availability of spectrum space for radar. External noise levels that can enter the radar via the antenna are higher at VHF than at microwave frequencies. Perhaps the chief limitation of operating radars at VHF is the difficulty of obtaining suitable spectrum space at these crowded frequencies.

In spite of its limitations, the VHF air surveillance radar was widely used by the Soviet Union because it was a large country, and the lower cost of VHF radars made them attractive for providing air surveillance over the large expanse of that country.[8] They have said they produced a large number of VHF air-surveillance radars—some were of very large size and long range, and most were readily transportable. It is interesting to note that VHF airborne intercept radars were widely used by the Germans in World War II. For example, the Lichtenstein SN-2 airborne radar operated from about 60 to over 100 MHz in various models. Radars at such frequencies were not affected by the countermeasure called *chaff* (also known as *window*).

UHF (300 to 1000 MHz). Many of the characteristics of radar operating in the VHF region also apply to some extent at UHF. UHF is a good frequency for Airborne Moving Target Indication (AMTI) radar in an Airborne Early Warning Radar (AEW), as discussed in Chapter 3. It is also a good frequency for the operation of long-range

radars for the detection and tracking of satellites and ballistic missiles. At the upper portion of this band there can be found long-range shipboard air-surveillance radars and radars (called *wind profilers*) that measure the speed and direction of the wind.

Ground Penetrating Radar (GPR), discussed in Chapter 21, is an example of what is called an ultrawideband (UWB) radar. Its wide signal bandwidth sometimes covers both the VHF and UHF bands. Such a radar's signal bandwidth might extend, for instance, from 50 to 500 MHz. A wide bandwidth is needed in order to obtain good range resolution. The lower frequencies are needed to allow the propagation of radar energy into the ground. (Even so, the loss in propagating through typical soil is so high that the ranges of a simple mobile GPR might be only a few meters.) Such ranges are suitable for locating buried power lines and pipe lines, as well as buried objects. If a radar is to see targets located on the surface but within foliage, similar frequencies are needed as for the GPR.

L band (1.0 to 2.0 GHz). This is the preferred frequency band for the operation of long-range (out to 200 nmi) air-surveillance radars. The Air Route Surveillance Radar (ARSR) used for long range air-traffic control is a good example. As one goes up in frequency, the effect of rain on performance begins to become significant, so the radar designer might have to worry about reducing the effect of rain at L-band and higher frequencies. This frequency band has also been attractive for the long-range detection of satellites and defense against intercontinental ballistic missiles.

S band (2.0 to 4.0 GHz). The Airport Surveillance Radar (ASR) that monitors air traffic within the region of an airport is at S band. Its range is typically 50 to 60 nmi. If a 3D radar is wanted (one that determines range, azimuth angle, and elevation angle), it can be achieved at S band.

It was said previously that long-range surveillance is better performed at low frequencies and the accurate measurement of target location is better performed at high frequencies. If only a single radar operating within a single frequency band can be used, then S band is a good compromise. It is also sometimes acceptable to use C band as the choice for a radar that performs both functions. The AWACS airborne air-surveillance radar also operates at S band. Usually, most radar applications are best operated in a particular frequency band at which the radar's performance is optimum. However, in the example of airborne air-surveillance radars, AWACS is found at S band and the U.S. Navy's E2 AEW radar at UHF. In spite of such a difference in frequency, it has been said that both radars have comparable performance.[9] (This is an exception to the observation about there being an optimum frequency band for each application.)

The Nexrad weather radar operates at S band. It is a good frequency for the observation of weather because a lower frequency would produce a much weaker radar echo signal from rain (since the radar echo from rain varies as the fourth power of the frequency), and a higher frequency would produce attenuation of the signal as it propagates through the rain and would not allow an accurate measurement of rainfall rate. There are weather radars at higher frequencies, but these are usually of shorter range than Nexrad and might be used for a more specific weather radar application than the accurate meteorological measurements provided by Nexrad.

C band (4.0 to 8.0 GHz). This band lies between S and X bands and has properties in between the two. Often, either S or X band might be preferred to the use of C band, although there have been important applications in the past for C band.

X band (8 to 12.0 GHz). This is a relatively popular radar band for military applications. It is widely used in military airborne radars for performing the roles of interceptor, fighter, and attack (of ground targets), as discussed in Chapter 5. It is also popular for imaging radars based on SAR and ISAR. X band is a suitable frequency for civil marine radars, airborne weather avoidance radar, airborne doppler navigation radars, and the police speed meter. Missile guidance systems are sometimes at X band. Radars at X band are generally of a convenient size and are, therefore, of interest for applications where mobility and light weight are important and very long range is not a major requirement. The relatively wide range of frequencies available at X band and the ability to obtain narrow beamwidths with relatively small antennas in this band are important considerations for high-resolution applications. Because of the high frequency of X band, rain can sometimes be a serious factor in reducing the performance of X-band systems.

K_u, K, and K_a Bands (12.0 to 40 GHz). As one goes to higher radar frequency, the physical size of antennas decrease, and in general, it is more difficult to generate large transmitter power. Thus, the range performance of radars at frequencies above X band is generally less than that of X band. Military airborne radars are found at K_u band as well as at X band. These frequency bands are attractive when a radar of smaller size has to be used for an application not requiring long range. The Airport Surface Detection Equipment (ASDE) generally found on top of the control tower at major airports has been at K_u band, primarily because of its better resolution than X band. In the original K band, there is a water-vapor absorption line at 22.2 GHz, which causes attenuation that can be a serious problem in some applications. This was discovered after the development of K-band radars began during World War II, which is why both K_u and K_a bands were later introduced. The radar echo from rain can limit the capability of radars at these frequencies.

Millimeter Wave Radar. Although this frequency region is of large extent, most of the interest in millimeter wave radar has been in the vicinity of 94 GHz where there is a minimum (called a *window*) in the atmospheric attenuation. (A window is a region of low attenuation relative to adjacent frequencies. The window at 94 GHz is about as wide as the entire microwave spectrum.) As mentioned previously, for radar purposes, the millimeter wave region, in practice, generally starts at 40 GHz or even at higher frequencies. The technology of millimeter wave radars and the propagation effects of the environment are not only different from microwave radars, but they are usually much more restricting. Unlike what is experienced at microwaves, the millimeter radar signal can be highly attenuated even when propagating in the clear atmosphere. Attenuation varies over the millimeter wave region. The attenuation in the 94 GHz window is actually higher than the attenuation of the atmospheric water-vapor absorption line at 22.2 GHz. The one-way attenuation in the oxygen absorption line at 60 GHz is about 12 dB per km, which essentially precludes its application. Attenuation in rain can also be a limitation in the millimeter wave region.

Interest in millimeter radar has been mainly because of its challenges as a frontier to be explored and put to productive use. Its good features are that it is a great place for employing wide bandwidth signals (there is plenty of spectrum space); radars can have high range-resolution and narrow beamwidths with small antennas; hostile electronic countermeasures to military radars are difficult to employ; and it is easier to have

a military radar with low probability of intercept at these frequencies than at lower frequencies. In the past, millimeter wave transmitters were not capable of an average power more than a few hundred watts—and were usually much less. Advances in gyrotrons (Chapter 10) can produce average power many orders of magnitude greater than more conventional millimeter-wave power sources. Thus, availability of high power is not a limitation as it once was.

Laser Radar. Lasers can produce usable power at optical frequencies and in the infrared region of the spectrum. They can utilize wide bandwidth (very short pulses) and can have very narrow beamwidths. Antenna apertures, however, are much smaller than at microwaves. Attenuation in the atmosphere and rain is very high, and performance in bad weather is quite limited. Receiver noise is determined by quantum effects rather than thermal noise. For several reasons, laser radar has had only limited application.

1.7 RADAR NOMENCLATURE

Military electronic equipment, including radar, is identified by the Joint Electronics Type Designation System (JETDS), as described in U.S. Military Standard MIL-STD-196D. The letter portion of the designation consists of the letters *AN*, a slant bar, and three additional letters appropriately selected to indicate where the equipment is installed, the type of equipment, and its purpose. Following the three letters are a dash and a numeral. The numeral is assigned in sequence for that particular combination of letters. Table 1.2 shows the letters that have been used for radar designations.

A suffix letter (*A*, *B*, *C*,…) follows the original designation for each modification of the equipment where interchangeability has been maintained. The letter *V* in parentheses added to the designation indicates variable systems (those whose functions may be varied through the addition or deletion of sets, groups, units, or combinations thereof). When the designation is followed by a dash, the letter *T*, and a number, the equipment is designed for training. In addition to the United States, these designations can also be used by Canada, Australia, New Zealand, and the United Kingdom. Special blocks of numbers are reserved for these countries. Further information can be found on the Internet under MIL-STD-196D.

The U.S. Federal Aviation Agency (FAA) uses the following to designate their air-traffic control radars:

- **ASR** Airport Surveillance Radar
- **ARSR** Air Route Surveillance Radar
- **ASDE** Airport Surface Detection Equipment
- **TDWR** Terminal Doppler Weather Radar

The numeral following the letter designation indicates the particular radar model (in sequence).

Weather radars developed by the U. S. Weather Service (NOAA) employ the designation WSR. The number following the designation is the year the radar went into service. Thus WSR-88D is the Nexrad doppler radar that first entered service in 1988. The letter *D* indicates it is a doppler weather radar.

TABLE 1.2 JETDS Letter Designations that Pertain to Radar

Installation (first letter)	Type of Equipment (second letter)	Purpose (third letter)
A. Piloted aircraft	**L**. Countermeasures	**B**. Bombing
B. Underwater mobile, submarine	**P**. Radar	**D**. Direction finder, reconnaissance and surveillance
D. Pilotless carrier	**S**. Special or combination	**G**. Fire control
F. Fixed ground	**W**. Armament (peculiar to armament not otherwise covered)	**N**. Navigation
G. General ground use		**Q**. Special or combination
K. Amphibious		**R**. Receiving
M. Mobile (ground)		**S**. Detecting/range and bearing, search
P. Portable		**T**. Transmitting
S. Water (ship)		**W**. Automatic flight or remote control
T. Transportable (ground)		**X**. Identification and recognition
U. General utility		**Y**. Surveillance and control (both fire control and air control)
V. Vehicular (ground)		
W. Water surface and underwater combined		
Z. Piloted-pilotless airborne vehicles combined		

1.8 SOME PAST ADVANCES IN RADAR

A brief listing of some of the major advances in technology and capability of radar in the twentieth century is given, in somewhat chronological but not exact order, as follows:

- The development of VHF radar for deployment on surface, ship, and aircraft for military air defense prior to and during World War II.
- The invention of the microwave magnetron and the application of waveguide technology early in WWII to obtain radars that could operate at microwave frequencies so that smaller and more mobile radars could be employed.
- The more than 100 different radar models developed at the MIT Radiation Laboratory in its five years of existence during WWII that provided the foundation for microwave radar.
- Marcum's theory of radar detection.
- The invention and development of the klystron and TWT amplifier tubes that provided high power with good stability.

- The use of the doppler frequency shift to detect moving targets in the presence of much larger echoes from clutter.
- The development of radars suitable for air-traffic control.
- Pulse compression.
- Monopulse tracking radar with good tracking accuracy and better resistance to electronic countermeasures than prior tracking radars.
- Synthetic aperture radar, which provided images of the ground and what is on it.
- Airborne MTI (AMTI) for long-range airborne air surveillance in the presence of clutter.
- Stable components and subsystems and ultralow sidelobe antennas that allowed high-PRF pulse doppler radar (AWACS) with large rejection of unwanted clutter.
- HF over-the-horizon radar that extended the range of detection of aircraft and ships by an order of magnitude.
- Digital processing, which has had a very major effect on improving radar capabilities ever since the early 1970s.
- Automatic detection and tracking for surveillance radars.
- Serial production of electronically scanned phased array radars.
- Inverse synthetic aperture radar (ISAR) that provided an image of a target as needed for noncooperative target recognition of ships.
- Doppler weather radar.
- Space radars suitable for the observation of planets such as Venus.
- Accurate computer calculation of the radar cross section of complex targets.
- Multifunction airborne military radar that are relatively small and lightweight that fit in the nose of a fighter aircraft and can perform a large number of different air-to-air and air-to-ground functions.

It is always a matter of opinion what the major advances in radar have been. Others might have a different list. Not every major radar accomplishment has been included in this listing. It could have been much longer and could have included multiple examples from each of the other chapters in this book, but this listing is sufficient to indicate the type of advances that have been important for improved radar capabilities.

1.9 APPLICATIONS OF RADAR

Military Applications. Radar was invented in the 1930s because of the need for defense against heavy military bomber aircraft. The military need for radar has probably been its most important application and the source of most of its major developments, including those for civilian purposes.

The chief use of military radar has been for air defense operating from land, sea, or air. It has not been practical to perform successful air defense without radar. In air defense, radar is used for long-range air surveillance, short-range detection of low-altitude "pop-up" targets, weapon control, missile guidance, noncooperative target recognition, and battle damage assessment. The proximity fuze in many weapons is

also an example of a radar. An excellent measure of the success of radar for military air defense is the large amounts of money that have been spent on methods to counter its effectiveness. These include electronic countermeasures and other aspects of electronic warfare, antiradiation missiles to home on radar signals, and low cross-section aircraft and ships. Radar is also used by the military for reconnaissance, targeting over land or sea, as well as surveillance over the sea.

On the battlefield, radar is asked to perform the functions of air surveillance (including surveillance of aircraft, helicopters, missiles, and unmanned airborne vehicles), control of weapons to an air intercept, hostile weapons location (mortars, artillery, and rockets), detection of intruding personnel, and control of air traffic.

The use of radar for ballistic missile defense has been of interest ever since the threat of ballistic missiles arose in the late 1950s. The longer ranges, high supersonic speeds, and the smaller target size of ballistic missiles make the problem challenging. There is no natural clutter problem in space as there is for defense against aircraft, but ballistic missiles can appear in the presence of a large number of extraneous confusion targets and other countermeasures that an attacker can launch to accompany the reentry vehicle carrying a warhead. The basic ballistic missile defense problem becomes more of a target recognition problem rather than detection and tracking. The need for warning of the approach of ballistic missiles has resulted in a number of different types of radars for performing such a function. Similarly, radars have been deployed that are capable of detecting and tracking satellites.

A related task for radar that is not military is the detection and interception of drug traffic. There are several types of radars that can contribute to this need, including the long-range HF over-the-horizon radar.

Remote Sensing of the Environment. The major application in this category has been weather observation radar such as the Nexrad system whose output is often seen on the television weather report. There also exist vertical-looking wind-profiler radars that determine wind speed and direction as a function of altitude, by detecting the very weak radar echo from the clear air. Located around airports are the Terminal Doppler Weather Radar (TDWR) systems that warn of dangerous wind shear produced by the weather effect known as the *downburst*, which can accompany severe storms. There is usually a specially designed weather avoidance radar in the nose of small as well as large aircraft to warn of dangerous or uncomfortable weather in flight.

Another successful remote-sensing radar was the downward-looking spaceborne altimeter radar that measured worldwide the *geoid* (the mean sea level, which is not the same all over the world), with exceptionally high accuracy. There have been attempts in the past to use radar for determining soil moisture and for assessing the status of agriculture crops, but these have not provided sufficient accuracy. Imaging radars in satellites or aircraft have been used to help ships efficiently navigate northern seas coated with ice because radar can tell which types of ice are easier for a ship to penetrate.

Air-Traffic Control. The high degree of safety in modern air travel is due in part to the successful applications of radar for the effective, efficient, and safe control of air traffic. Major airports employ an Airport Surveillance Radar (ASR) for observing the air traffic in the vicinity of the airport. Such radars also provide information about nearby weather so aircraft can be routed around uncomfortable weather. Major airports also have a radar called Airport Surface Detection Equipment (ASDE) for observing

and safely controlling aircraft and airport vehicle traffic on the ground. For control of air traffic en route from one terminal to another, long-range Air Route Surveillance Radars (ARSR) are found worldwide. The Air Traffic Control Radar Beacon System (ATCRBS) is not a radar but is a cooperative system used to identify aircraft in flight. It uses radar-like technology and was originally based on the military IFF (Identification Friend or Foe) system.

Other Applications. A highly significant application of radar that provided information not available by any other method, was the exploration of the surface of the planet Venus by an imaging radar that could see under the ever-present clouds that mask the planet. One of the widest used and least expensive of radars has been the civil marine radar found throughout the world for the safe navigation of boats and ships. Some readers have undoubtedly been confronted by the highway police using the CW doppler radar to measure the speed of a vehicle. Ground penetrating radar has been used to find buried utility lines, as well as by the police for locating buried objects and bodies. Archeologists have used it to determine where to begin to look for buried artifacts. Radar has been helpful to both the ornithologist and entomologist for better understanding the movements of birds and insects. It has also been demonstrated that radar can detect the gas seepage that is often found over underground oil and gas deposits.[10]

1.10 CONCEPTUAL RADAR SYSTEM DESIGN

There are various aspects to radar system design. But before a new radar that has not existed previously can be manufactured, a *conceptual design* has to be performed to guide the actual development. A conceptual design is based on the requirements for the radar that will satisfy the customer or user of the radar. The result of a conceptual design effort is to provide a list of the radar characteristics as found in the radar equation and related equations and the general characteristics of the subsystems (transmitter, antenna, receiver, signal processing, and so forth) that might be employed. The radar equation is used as an important guide for determining the various tradeoffs and options available to the radar system designer so as to determine a suitable concept to meet the desired need. This section briefly summarizes how a radar systems engineer might begin to approach the conceptual design of a new radar. There are no firmly established procedures to carry out a conceptual design. Every radar company and every radar design engineer develops his or her own style. What is described here is a brief summary of one approach to conceptual radar design.

General Guideline. It should be mentioned that there are at least two ways by which a new radar system might be produced for some particular radar application. One method is based on exploiting the advantages of some new invention, new technique, new device, or new knowledge. The invention of the microwave magnetron early in World War II is an example. After the magnetron appeared, radar design was different from what it had been before. The other, and probably more common method for conceptual radar system design, is to start with what the new radar has to do, examine the various approaches available to achieve the desired capability, carefully evaluate each approach, and then select the one that best meets the needs within the operational and fiscal constraints imposed. In brief, it might consist of the following steps:

- *Description of the need or problem to be solved.*
  This is from the viewpoint of the customer or the user of the radar.
- *Interaction between the customer and the systems engineer.*
  This is for the purpose of exploring the tradeoffs, which the customer might not be aware of, that might allow the customer to better obtain what is wanted without excessive cost or risk. Unfortunately, interaction between the potential user of the radar and the radar systems engineer is not always done in competitive procurements.
- *Identification and exploration of possible solutions.*
  This includes understanding the advantages and limitations of the various possible solutions.
- *Selection of the optimum or near optimum solution.*
  In many engineering endeavors, optimum does not mean the best since the best might not be affordable or achievable in the required time. *Optimum*, as used here, means the best under a given set of assumptions. Engineering often involves achieving a near-optimum, not the optimum. Selecting the preferred solution should be based on a well-defined criterion.
- *Detailed description of the selected approach.*
  This is in terms of the characteristics of the radar and the type of subsystems to be employed.
- *Analysis and evaluation of the proposed design.*
  This is to verify the correctness of the selected approach.

As one proceeds through this process, one might reach a "dead end" and have to start over—sometimes more than once. Having to start over is not unusual during a new design effort.

One cannot devise a unique set of guidelines for performing the design of a radar. If that were possible, radar design could be done entirely by computer. Because of the usual lack of complete information, most engineering design requires, at some point, the judgment and experience of the design engineer in order to succeed.

The Radar Equation in Conceptual Design. The radar equation is the basis for conceptual radar system design. Some parameters of the radar equation are determined by what the radar is required to do. Others may be decided upon unilaterally by the customer—but that should be done with caution. The customer usually should be the one who states the nature of the radar target, the environment in which the radar is to operate, restrictions on size and weight, the use to which the radar information is to be put, and any other constraints that have to be imposed. From this information, the radar systems engineer determines what is the radar cross section of the target, the range and angle accuracies needed to meet the radar user's needs, as well as the antenna revisit time. Some parameters, such as antenna gain, might be affected by more than one need or requirement. For instance, a particular antenna beamwidth might be influenced by the tracking accuracy, resolution of nearby targets, the maximum size the antenna can be for a particular application, the need for a desired radar range, and the choice of radar frequency. The radar frequency is usually affected by many things, including the availability of allowed frequencies at which to operate. The radar frequency might be the last parameter of the radar to be selected—after many other compromises have been made.

REFERENCES

1. *IEEE Standard Dictionary of Electrical and Electronic Terms*, 4th Ed. New York: IEEE, 1988.
2. M. I. Skolnik, G. Linde, and K. Meads, "Senrad: an advanced wideband air-surveillance radar," *IEEE Trans*., vol. AES-37, pp. 1163–1175, October 2001.
3. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill, 2001, Fig. 2.6.
4. F. E. Nathanson, *Radar Design Principles*, New York: McGraw-Hill, 1991, Fig. 2.2.
5. This table has been derived from IEEE Standard Letter Designations for Radar-Frequency Bands, IEEE Std. 521-2002.
6. Specific radiolocation frequency ranges may be found in the "FCC Online Table of Frequency Allocations," 47 C.F.R. § 2.106.
7. "Performing electronic countermeasures in the United States and Canada," U.S. Navy OPNAVINST 3430.9B, October 27, 1969. Similar versions issued by the U.S. Air Force, AFR 55-44; U.S. Army, AR 105-86; and U.S. Marine Corps, MCO 3430.1.
8. A. Zachepitsky, "VHF (metric band) radars from Nizhny Novgorod Research Radiotechnical Institute," *IEEE AES Systems Magazine*, vol. 15, pp. 9–14, June 2000.
9. Anonymous, "AWACS vs. E2C battle a standoff," *EW Magazine*, p. 31, May/June 1976.
10. M. Skolnik, D. Hemenway, and J. P. Hansen, "Radar detection of gas seepage associated with oil and gas deposits," *IEEE Trans*, vol. GRS-30, pp. 630–633, May 1992.

Chapter 2
MTI Radar

William W. Shrader

Shrader Associates, Inc.

Vilhelm Gregers-Hansen

Naval Research Laboratory

2.1 PREFACE

This chapter addresses surface-based radars; e.g., radars sited on land or installed onboard ships. For airborne radar, rapid platform motion has a significant effect on design and performance as discussed in Chapters 3, 4, and 5 of this *Handbook.*

The fundamental theory of moving target indication (MTI) radar, as presented in the previous editions of the *Radar Handbook*, has not materially changed. The performance of MTI radar, however, has been greatly improved, due primarily to four advances: (1) increased stability of radar subsystems such as transmitters, oscillators, and receivers; (2) increased dynamic range of receivers and analog-to-digital converters (A/D); (3) faster and more powerful digital processing; and (4) better awareness of the limitations, and therefore requisite solutions, of adapting MTI systems to the environment. These four advances have made it practical to use sophisticated techniques that were considered, and sometimes tried, many years ago but were impractical to implement. Examples of early concepts that were well ahead of the available technology were the velocity indicating coherent integrator (VICI)[1] and the coherent memory filter (CMF).[2,54]

Although these improvements have enabled much improved MTI capabilities, there are still no perfect solutions to all MTI radar problems, and the design of an MTI system is still as much of an art as it is a science. Examples of current problems include the fact that when receivers are built with increased dynamic range, system instability limitations will cause increased clutter residue (relative to system noise) that can cause false detections. Clutter maps, which are used to prevent false detections from clutter residue, work quite well on fixed radar systems, but are difficult to implement on, for example, shipboard radars, because as the ship moves, the aspect and range to each clutter patch changes, creating increased residues after the clutter map. A decrease in the resolution of the clutter map to counter the rapidly changing clutter residue will preclude much of the interclutter visibility (see later in this chapter), which is one of the least appreciated secrets of successful MTI operation.

MTI radar must work in the environment that contains strong fixed clutter, birds, bats and insects, weather, automobiles, and ducting. The ducting, also referred to as *anomalous propagation*, causes radar returns from clutter on the surface of the Earth to appear

at greatly extended ranges, which exacerbates the problems with birds and automobiles, and can also cause the detection of fixed clutter hundreds of kilometers away.

The clutter models contained in this chapter are approximations of the types of clutter that must be addressed. The exact quantitative data, such as precise spectrum and amplitude of each type of clutter, or the exact number of birds or point reflectors (e.g., water towers or oil-well derricks) per unit area, is not important, because the MTI radar designer must create a robust system that will function well no matter the actual deviation from the clutter models of real clutter encountered.

MTI radars may use rotating antennas or fixed apertures with electronic beam scanning (phased arrays). The rotating antenna may use a continuous waveform processed through either a finite-impulse-response (FIR) filter or an infinite-impulse-response (IIR) filter, or may use a batch waveform consisting of coherent processing intervals (CPIs) that are processed in FIR filters in groups of N pulses. (The term *MTI filter*, used often in this chapter, is a generic nomenclature that includes both FIR and IIR filters.) The finite time-on-target dictates the need for a batch processing approach.

There are many different combinations of successful MTI techniques, but any specific MTI radar system must be a total concept based on the parameters of the antenna, transmitter, waveform, signal processing, and the operational environment.

A number of plan-position-indicator (PPI) photographs, taken years ago, are included in this chapter to provide a better understanding of the environment that is difficult to appreciate with many modern radars. These photographs show MTI operation, birds, insects, and ducting better than can be described in words.

Attention is especially directed to the final section in this chapter, "Considerations Applicable to MTI Radar Systems," which provides insight into both hardware and environmental lessons learned during many decades of MTI system development.

2.2 INTRODUCTION TO MTI RADAR

The purpose of MTI radar is to reject returns from fixed or slow-moving unwanted targets, such as buildings, hills, trees, sea, and rain, and retain for detection or display signals from moving targets such as aircraft. Figure 2.1 shows a pair of photographs of a PPI, which illustrates the effectiveness of such an MTI system. The distance from the center to the edge of the PPI is 40 nmi. The range marks are at 10-nmi intervals. The picture on the left is the normal video display, showing mainly the fixed-target returns. The picture on the right shows the effectiveness of the MTI clutter rejection. The camera shutter was left open for three scans of the antenna; thus, aircraft show up as a succession of three returns. MTI radar utilizes the doppler shift imparted on the reflected signal by a moving target to distinguish moving targets from fixed targets. In a pulse radar system, this doppler shift appears as a change of phase of received signals between consecutive radar pulses. Consider a radar that transmits a pulse of radio frequency (RF) energy that is reflected by both a building (fixed target) and an airplane (moving target) approaching the radar. The reflected pulses return to the radar a certain time later. The radar then transmits a second pulse. The reflection from the building occurs in exactly the same amount of time, but the reflection from the moving aircraft occurs in less time because the aircraft has moved closer to the radar in the interval between transmitted pulses. The precise time that it takes the reflected signal to reach the radar is not of fundamental importance. What is significant is whether the time changes between pulses. The time change, which is of the order of a few nanoseconds for an aircraft target, is determined by comparing the phase of the received signal with

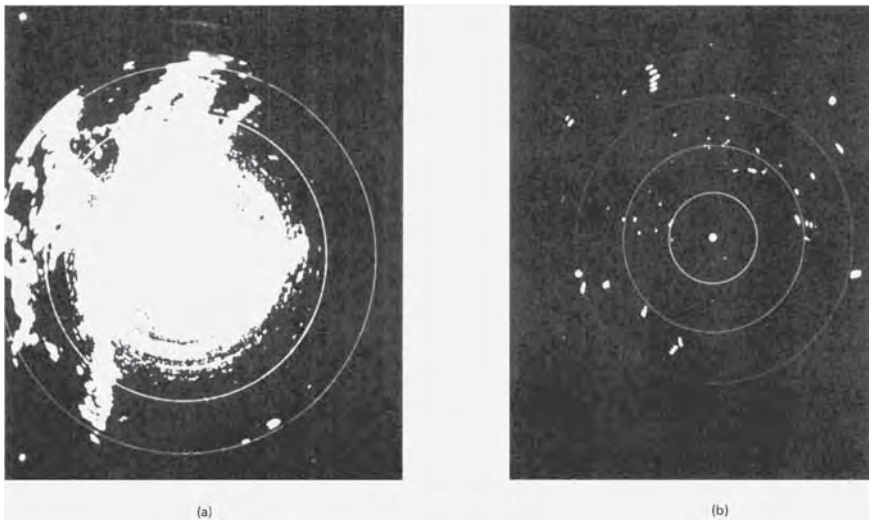


FIGURE 2.1 (*a*) Normal video and (*b*) MTI video: These PPI photographs show how effective an MTI system can be. Aircraft appear as three consecutive blips in the right-hand picture because the camera shutter was open for three revolutions of the antenna. The PPI range is 40 nmi.

the phase of a reference oscillator in the radar. If the target moves between pulses, the phase of the received pulse changes.

Figure 2.2 shows a simplified block diagram of a coherent MTI system. The RF oscillator feeds the pulsed amplifier, which transmits the pulses. The RF oscillator

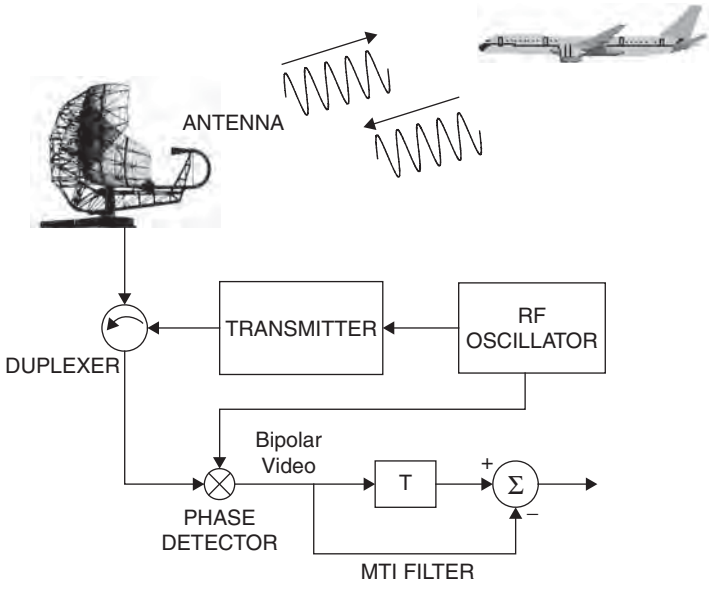


FIGURE 2.2 Simplified block diagram of a coherent MTI system

is also used as a phase reference for determining the phase of reflected signals. The phase information is stored in a pulse repetition interval (PRI)) memory for the period, *T,* between transmitted pulses, and is subtracted from the phase information from the current received pulse. There is an output from the subtractor only when a reflection has occurred from a moving target.

Moving-Target Indicator (MTI) Block Diagram. A more complete block diagram of an MTI radar is shown in Figure 2.3. This block diagram is representative of a modern air traffic control radar operating at L or S band with a typical interpulse period of 1–3 ms and a CW pulse length of a few μs when the transmitter employs a vacuum tube amplifier such as, for example, a klystron, or tens of μs for a pulse compression waveform when a solid-state transmitter is used. The received signals are amplified in a low-noise amplifier (LNA) and subsequently downconverted through one or more intermediate frequencies (IF) by mixing with stable local oscillators. A bandpass IF limiter at the receiver output protects the A/D converter from damage but also prevents limiting from taking place in the A/D converter. In early MTI systems, the IF limiter served the purpose of deliberately restricting the dynamic range to reduce clutter residues at the MTI output. The received signals are then converted into in-phase and quadrature components (I & Q) through the A/D converter, either using a pair of phase detectors or through direct sampling as discussed in Section 2.13. The in-phase (I) and quadrature (Q) outputs are a function of the amplitude and phase of the IF signal and

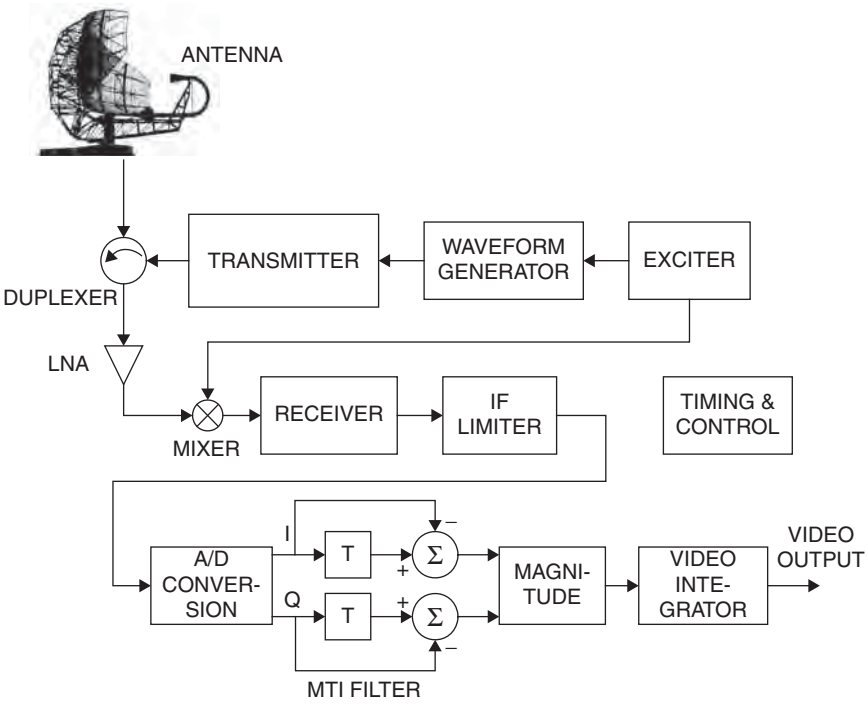


FIGURE 2.3 MTI system block diagram

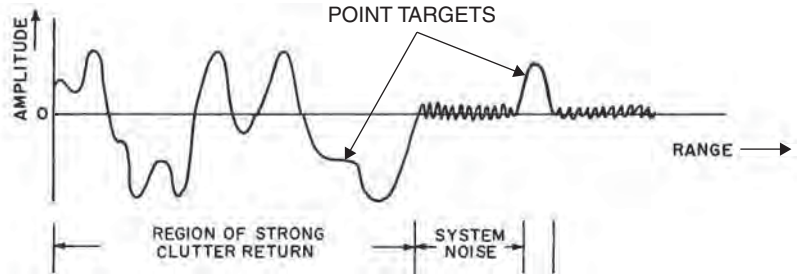


FIGURE 2. 4 Bipolar video: return from single transmitter pulse

have in the past been referred to as bipolar videos, but a more correct terminology is that of the complex envelope of the received signals. An example of such a bipolar video (either I or Q), received from a single transmitted pulse and including both clutter and point targets is sketched in Figure 2.4. If the point targets are moving, the superimposed bipolar video from several transmitted pulses would appear as in Figure 2.5.

The remainder of the block diagram in Figure 2.3 shows the remaining processing required so that the moving targets can be displayed on a PPI or sent to an automatic target extractor. The in-phase and quadrature outputs from the A/D converter are stored in a PRI memory and also subtracted from the output from the previous transmitted pulse. This implementation represents the most basic two-pulse MTI canceler implemented as a finite impulse response (FIR) filter. As discussed in Section 2.8, MTI cancelers used in practical radars use higher-order filters, and these are sometimes implemented as infinite impulse response (IIR) filters.

The output of the subtractors is again a bipolar signal that contains moving targets, system noise, and a small amount of clutter residue if the clutter cancellation is not perfect. The magnitudes of the in-phase and quadrature signals are then computed ($\sqrt{I^2+Q^2}$) and converted to analog video in a digital-to-analog (D/A) converter for display on a PPI. The digital signal may also be sent to automatic target detection circuitry. The dynamic range (peak signal to rms noise) is limited to about 20 dB for a PPI display.

A key distinction, sometimes lost in the complexities of the systems that follow, is that an MTI radar system eliminates fixed clutter because the phase of signals returned from consecutive transmitted pulses do not (appreciably) change. The fixed clutter is removed after as few as two transmitted pulses by the subtraction process described

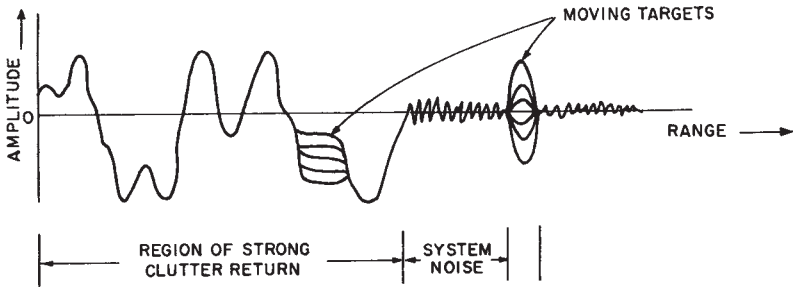


FIGURE 2.5 Bipolar video: from consecutive transmitted pulses

above, even if each transmitted pulse has frequency modulation or other artifacts, as long as the artifacts are identical pulse-to-pulse. The point being made here is that MTI system operation does not depend on the frequency resolution of targets from clutter. To provide frequency resolution would require much longer dwell times on target than two pulses separated by a single PRI. Such extended dwell times is one of the fundamental characteristics of the moving target detector.

Moving-Target Detector (MTD) Block Diagram. Progress in digital signal processing technology by the mid-1970s made it practical for the first time to improve the performance of the classical MTI by (1) implementing a parallel bank of FIR filters to increase the output signal-to-clutter ratio and (2) replacing the IF limiter used in the past with a high-resolution clutter map for effective false alarm control. Although these concepts had been explored many years earlier using the Velocity Indicating Coherent Integrator (VICI)[1] or the Coherent Memory Filter (CMF)[2,54] to implement a doppler filter bank, and storage tubes or magnetic drum memory to implement clutter maps, it was the work at the MIT Lincoln Laboratory to improve the performance of airport surveillance radars that resulted in one of the first working examples of what has become known as the Moving Target Detection (MTD) radar.[3,4] The theory and expected benefits of this approach were described in two reports in 1972,[5] which provided the mathematical foundation for the understanding and the practical implementation of the MTD concept.

The predicted subclutter visibility improvement for the ASR-7 airport surveillance radar, when the three-pulse MTI processor was replaced by the second-generation MTD II processor, is shown in Figure 2.6.

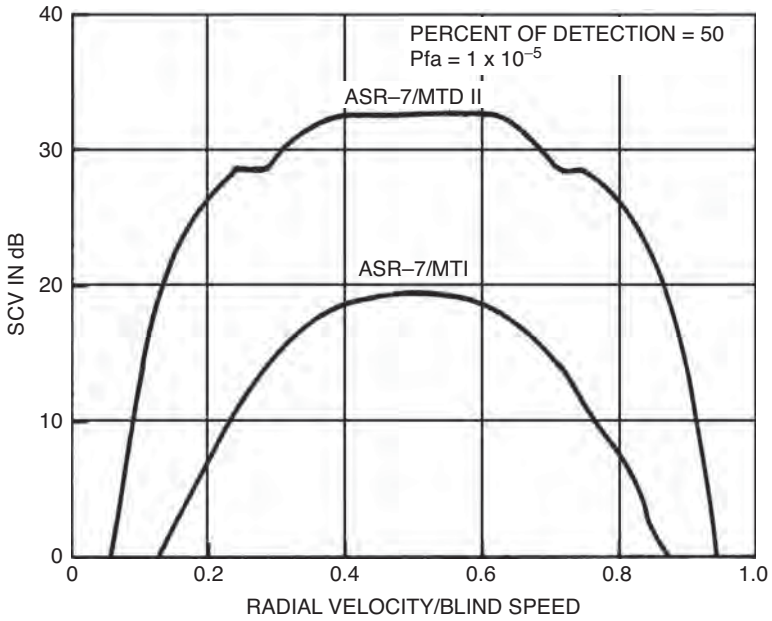


FIGURE 2.6 Subclutter visibility comparison between three-pulse MTI and MTD II

Part of this improvement was due to the use of doppler filter designs utilizing eight pulses, instead of just three for the MTI, and part was the result of allowing a larger dynamic range into the MTD processor and relying on a clutter map to suppress residues in regions where the clutter level exceeds the maximum clutter suppression of the radar.

The block diagram of the MTD II signal processor is shown in Figure 2.7. Parallel processing channels are provided for moving targets through the two-pulse MTI canceler and the seven-pulse doppler filter bank, and for nonmoving ("zero-doppler") targets through the 0-Velocity Filter. A high resolution clutter map is built from the "0-Velocity Filter" output, and the clutter map content is used for thresholding in the two processing channels. In the moving target channel, the threshold obtained from the clutter map content is scaled down by the expected clutter attenuation. In addition to the clutter map thresholding, conventional constant false alarm rate thresholding is utilized against moving clutter (rain) and interference. Detection outputs, named *Primitive Target Outputs*, are obtained through this processing for each individual processed CPI. Figure 2.8 shows the additional processing required to generate centroided Target Reports and the processing of these Target Reports to obtain track outputs for display to the air traffic control system.

The MTD radar transmits a group of N pulses at a constant pulse repetition frequency (PRF) and at a fixed radar frequency. This set of pulses is usually referred to as the *coherent processing interval (CPI)* or *pulse batch.* Sometimes one or two additional fill pulses are added to the CPI in order to suppress range-ambiguous clutter returns, as might occur during periods of anomalous propagation. The returns received during one CPI are processed in the bank of N-pulse finite-impulse-response (FIR) filters. Then the radar may change its PRF and/or RF frequency and transmit another CPI of N pulses. Since most search radars are ambiguous in doppler, the use of different

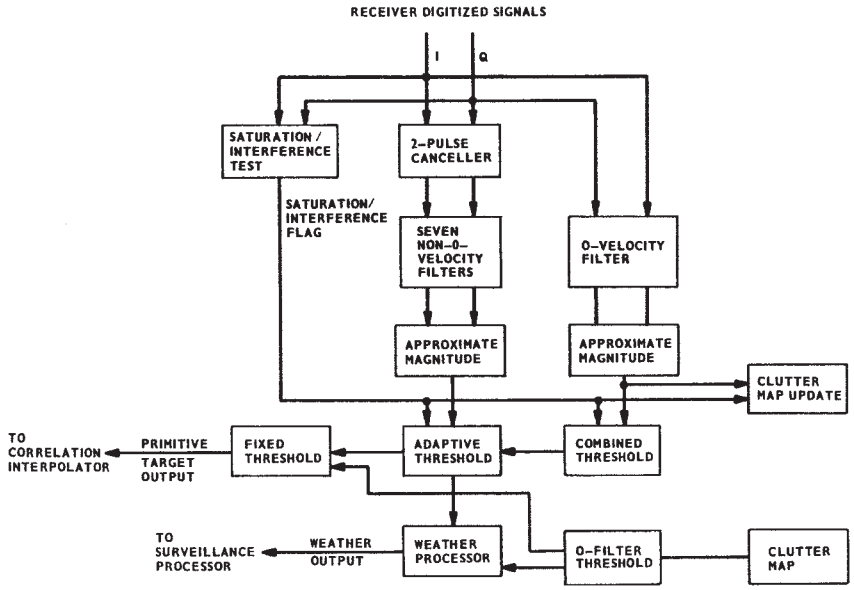


FIGURE 2.7 Block diagram of MTD II signal processor

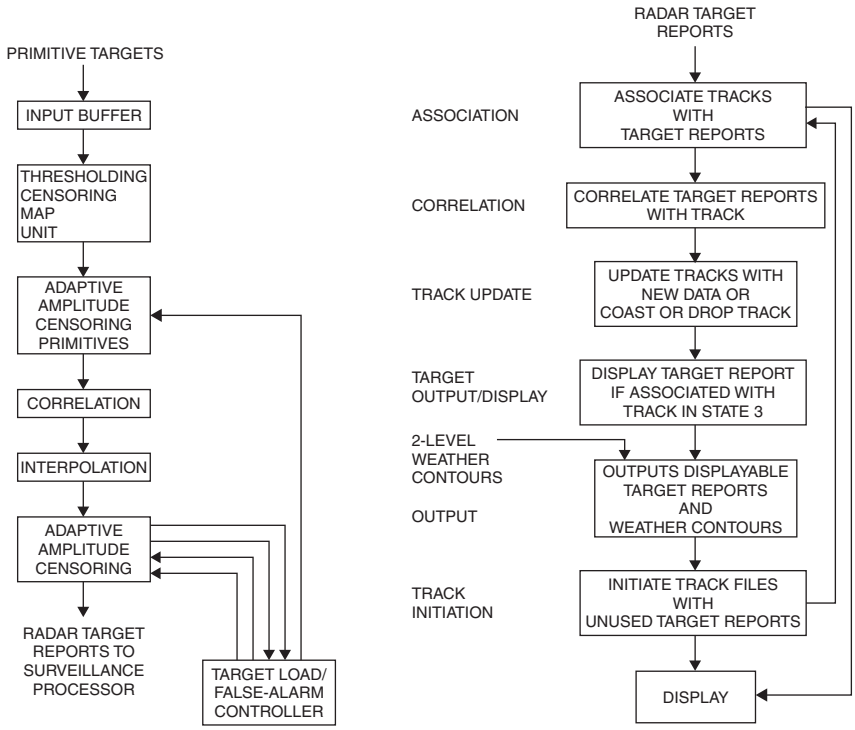


FIGURE 2.8 Processing of Primitive Target detections and Radar Target Reports in MTD II

PRFs on successive coherent dwells will cause the target response to fall at different frequencies of the filter passband on the successive opportunities during the time on target, thus eliminating blind speeds.

Each doppler filter is designed to respond to targets in nonoverlapping portions of the doppler frequency band and to suppress sources of clutter at all other doppler frequencies. This approach maximizes the coherent signal integration in each doppler filter and provides clutter attenuation over a larger range of doppler frequencies than achievable with a single MTI filter. Thus, one or more clutter filters may suppress multiple clutter sources located at different doppler frequencies. An example of the use of an MTD doppler filter bank against simultaneous land and weather clutter (Wx) is illustrated in Figure 2.9. It can be seen that filters 3 and 4 will provide significant suppression of both clutter sources.

The output of each doppler filter is envelope-detected and processed through a cell-averaging constant false alarm rate (CA-CFAR) processor to suppress residues due to range-extended clutter that may not have been fully suppressed by the filter.

As will be discussed later in this chapter, the conventional MTI detection system often relies on a carefully controlled dynamic range in the IF section of the radar receiver to ensure that clutter residues at the MTI output are suppressed to the level of the receiver noise or below. This limited dynamic range, however, has the undesirable effect of causing additional clutter spectral broadening, and the achievable clutter suppression is consequently reduced.

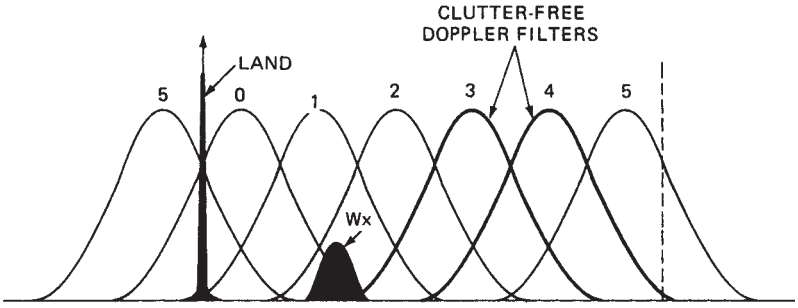


FIGURE 2.9 Suppression of multiple clutter sources by using a doppler filter bank

In the MTD, one or more high-resolution clutter maps are used to suppress the clutter residues, after doppler filtering, to the receiver noise level (or, alternatively, to raise the detection threshold above the level of the residues). This in turn eliminates the need to restrict the IF dynamic range, which can then be set to the maximum value supported by the A/D converters. Thus, a system concept is obtained that provides a clutter suppression capability that is limited only by the radar system stability, the dynamic range of the receiver-processor, and the spectrum width of the returns from clutter. The concept of a high-resolution digital clutter map to suppress clutter residues is related to earlier efforts to construct analog area MTI systems using, for example, storage tubes.

Also included in the MTD implementation are "…area thresholds maintained to control excessive false alarms, particularly from bird flocks. Each area of about 16 square nautical miles is divided into several velocity regions. The threshold in each region is adjusted on each scan to achieve the desired limit on false alarms without raising the threshold so high that small aircraft are prevented from being placed in track status."[4]

In subsequent sections, specific aspects of the design of an MTD system will be discussed. Thus, Section 2.10 will discuss the design and performance of doppler filter banks, and a detailed discussion of clutter maps will follow in Section 2.15. Since the original work at Lincoln Laboratory to develop the MTD concept, a number of MTD systems have been developed that vary in detail from the original concept. Also, the use of clutter maps to inhibit excessive clutter residue, instead of controlling clutter residue with intentionally restricted dynamic range, has been adopted in newer MTI systems.

2.3 CLUTTER FILTER RESPONSE TO MOVING TARGETS

The response of an MTI system to a moving target varies as a function of the target's radial velocity. For the MTI system described above, the response, normalized for unity noise power gain, is shown in Figure 2.10. Note that there is zero response to stationary targets and also to targets at ±89, ±178, ±267,. . . knots. These speeds, known as *blind speeds,* are where the targets move 0, ½, 1, 1½, . . . wavelengths between consecutive transmitted pulses. This results in the received signal being

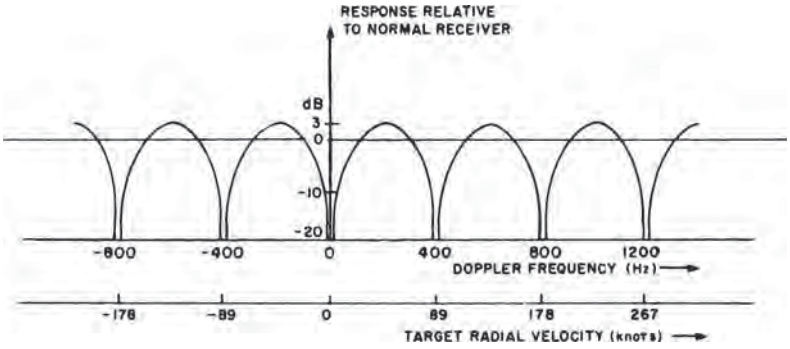


FIGURE 2.10 MTI system response for 1300-MHz radar operating at 400 pps

shifted precisely 360° or multiples thereof between pulses, which results in no change in the phase-detector output. The blind speeds can be calculated

$$V_B = k \cdot \frac{\lambda \cdot f_r}{2} \qquad k = \pm 0,\ 1,\ 2,... \tag{2.1}$$

where V_B is the blind speed, in meters per second; λ is the transmitted wavelength, in meters; and f_r is the PRF, in hertz. A convenient set of units for this equation is

$$V_B \text{ (knots)} = k \cdot \frac{0.29 \cdot f_r}{f_{\text{GHz}}} \qquad k = \pm 0,\ 1,\ 2,... \tag{2.2}$$

where f_r is the PRF (pulse repetition frequency) in hertz; and f_{GHz} is the transmitted frequency, in gigahertz. Note from the velocity response curve that the response to targets at velocities midway between the blind speeds is greater than the response for a normal receiver.

The abscissa of the velocity response curve can also be labeled in terms of doppler frequency. The doppler frequency of the target can be calculated from

$$f_d = \frac{2 \cdot V_R}{\lambda} \tag{2.3}$$

where f_d is the doppler frequency, in hertz; V_R is the target radial velocity, in meters per second; and λ is the transmitted wavelength, in meters. It can be seen from Figure 2.10 that the doppler frequencies for which the system is blind occur at multiples of the pulse repetition frequency.

2.4 CLUTTER CHARACTERISTICS

The clutter suppression needed from an MTI or MTD radar depends on the characteristics of the clutter environment, the specific radar target detection requirements, and the major radar design characteristics such as range and angle resolution as well as operating frequency. The ability of a radar to suppress clutter is determined by radar

waveform and processing, available dynamic range, and the overall radar system stability. In this section, some of the key characteristics of radar clutter and its influence on MTI radar design will be summarized.

Spectral Characteristics. The spectral characteristics of clutter, as discussed in most references, implicitly assumes that the radar transmits a continuous, constant PRF waveform. The spectrum of the output of a pulsed transmitter using a simple rectangular pulse of length τ is shown in Figure 2.11. The spectral width of the (sin U)/U envelope is determined by the transmitted pulse width, the first nulls occurring at a frequency of $f_0 \pm 1/\tau$. The individual spectral lines are separated by a frequency equal to the PRF. These spectral lines fall at precisely the same frequencies as the nulls of the MTI filter response shown in Figure 2.10. Thus, a canceler will, in theory, fully reject clutter with this ideal line spectrum. In practice, however, the spectral lines of the clutter returns are broadened by motion of the clutter (such as windblown trees or waves on the sea surface) as well as by the motion of the antenna in a scanning radar or due to platform motion. This spectral spread prevents perfect cancellation of clutter in an MTI system.

Often, in the past, the assumption has been made that the returns from clutter have a gaussian power spectral density, which may be characterized by its standard deviation σ_v and mean velocity m_v, both in units of m/s.[6] Using this gaussian model, each of the spectral lines in Figure 2.11 will be convolved with the spectrum:

$$S_G(f) = \frac{1}{\sqrt{2\pi}\sigma_f} \cdot \exp\left(-\frac{(f - m_f)^2}{2\sigma_f^2}\right) \tag{2.4}$$

This spectrum is normalized to have unit power, and the velocity parameters have been converted to Hz using the doppler equation:

$$\begin{aligned} m_f &= \frac{2 \cdot m_v}{\lambda} \\ \sigma_f &= \frac{2 \cdot \sigma_v}{\lambda} \end{aligned} \tag{2.5}$$

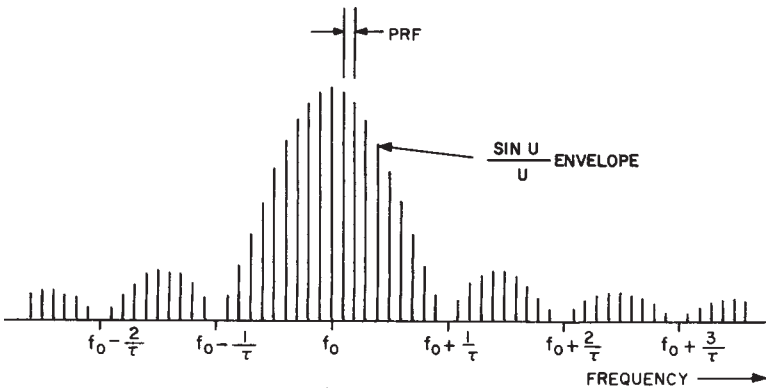


FIGURE 2.11 Pulse transmitter spectrum

where λ is the radar wavelength. Instead of the standard deviation σ_f, the power spectrum can be defined by its 3-dB width B_3, as follows:

$$S_G(f) = \frac{\sqrt{4 \cdot \ln(2)}}{\sqrt{\pi} \cdot B_3} \cdot \exp\left(-\frac{4 \cdot \ln(2) \cdot f^2}{B_3^2}\right) \tag{2.6}$$

where

$$B_3 = \sqrt{8 \cdot \ln(2)} \cdot \sigma_f = 2.3548 \cdot \sigma_f \tag{2.7}$$

The early experimental results that led to the general adoption of the gaussian model[6] were obtained with radar equipment of limited stability, and the spectral shape was sometimes derived from video spectra computed using square-law detected returns.

By the mid-1970s, new experimental results were obtained,[7,8] which showed that the spectrum fall-off was slower than predicted by the gaussian model. This led to new models based on polynomial representations of the spectrum using an equation of the form:

$$S_{\text{POLY}}(f) = \frac{n \cdot \sin\left(\frac{\pi}{n}\right)}{\pi \cdot B_3} \cdot \frac{1}{1 + \left(\frac{2\,|f|}{B_3}\right)^n} \tag{2.8}$$

The spectrum shape is determined by the integer n, which must be 4 or larger in order for the two first spectral moments to exist. A typical value used for this spectrum is $n = 4$ which results in

$$S_{\text{POLY}}(f) = \frac{\sqrt{8}}{\pi \cdot B_3} \cdot \frac{1}{1 + \left(\frac{2\,|f|}{B_3}\right)^4} \tag{2.9}$$

The relationship between the standard deviation of this spectrum and its 3-dB width is given by

$$B_3 = 2 \cdot \sigma_f \tag{2.10}$$

A potential issue with this model is that the skirts of the spectrum correspond to very large radial velocity components of the clutter internal motion.

During the 1990s, an extensive measurement program conducted at the MIT Lincoln Laboratory obtained more accurate data on land clutter spectra using a very stable radar equipment and data was collected under well-controlled conditions.[9] These new results led to the following exponential model for land clutter spectra:

$$S_{\text{EXP}}(f) = \frac{\ln(2)}{B_3} \cdot \exp\left(-\frac{2 \cdot \ln(2)}{B_3} \cdot |f|\right) \tag{2.11}$$

Here the 3-dB spectrum width can be expressed in terms of the standard deviation by

$$B_3 = \sqrt{2} \cdot \ln(2) \cdot \sigma_f = 0.9803 \cdot \sigma_f \tag{2.12}$$

Billingsley[9] used the parameters g, v_c, and β, respectively, for the gaussian, the polynomial, and the exponential spectrum models. In addition, the exponent n is needed for the polynomial model. These parameters were chosen to simplify the functional description of the spectrum shape. In terms of the standard deviation of the spectral width in m/s, these parameters can be defined as follows:

$$\begin{aligned} g &= \frac{1}{2\cdot\sigma_v^2} && \text{- gaussian spectrum} \\ v_c &= \sqrt{2\cdot\ln(2)}\cdot\sigma_v && \text{- polynomial spectrum with } n=4 \\ \beta &= \frac{\sqrt{2}}{\sigma_v} && \text{- exponential spectrum} \end{aligned} \tag{2.13}$$

Assuming a value of $\sigma_v = 0.25$ m/s, corresponding to windy condition, the three-clutter spectrum models are compared in Figure 2.12. As noted in Billingsley[9] all three models are in reasonable agreement for the upper 30–40 dB of their range but differ appreciably at the lower values of clutter spectral density.

Estimated values of the spectral spread of land clutter from forested regions and for different wind speeds are shown in Table 2.1. The values in the table are based on Billingsley's parameter β, but columns have been added with the corresponding rms spectral spread in m/s. An example of a measured land clutter spectrum is shown in Figure 2.13. The spectral shape parameter β can be estimated as the slope of the (upper) skirt of the spectrum in dB per m/s divided by 10/ln(10). These values of β were added in this figure.

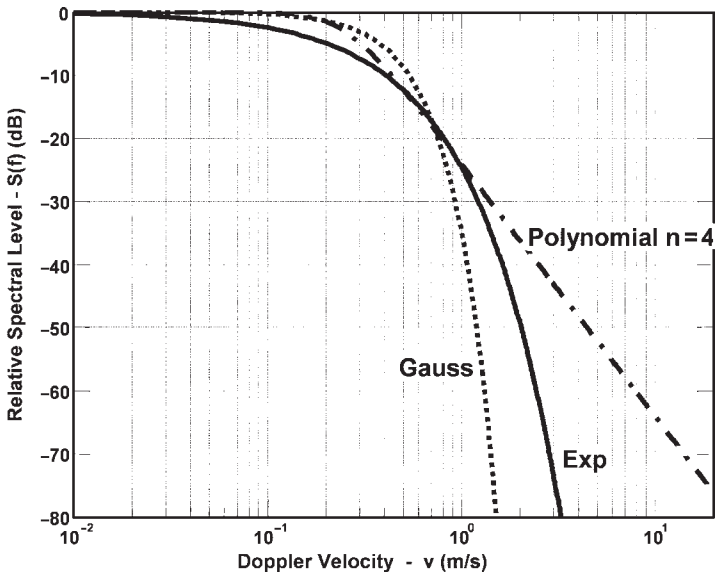


FIGURE 2.12 Comparison of gaussian, exponential, and polynomial spectra for an rms spectral spread of $\sigma_v = 0.25$ m/s

TABLE 2.1 Measured Spectral Spread for Different Wind Conditions (*after J.B.Billingsley*[9] *© William Andrew Publishing Inc. 2002*)

Wind Conditions	Wind Speed (mph)	Exponential ac Shape Parameter β $(m/s)^{-1}$		RMS Spectral Width σ_v (m/s)	
		Typical	Worst Case	Typical	Worst Case
Light air	1–7	12	—	0.12	—
Breezy	7–15	8	—	0.18	—
Windy	15–30	5.7	5.2	0.25	0.27
Gale force (est.)	30–60	4.3	3.8	0.33	0.37

The values of rms spectral spread of land clutter as derived from the data in Billingsley[9] agree quite well with previous studies. It can probably safely be stated that the polynomial model of land clutter spectra is far too pessimistic at spectral values below –40 dB and should be avoided for radar analysis requiring a large clutter attenuation value.

The case for the exponential model, as presented by Billingsley, is quite convincing, and this model has been widely accepted as being the most accurate for radar performance predictions.

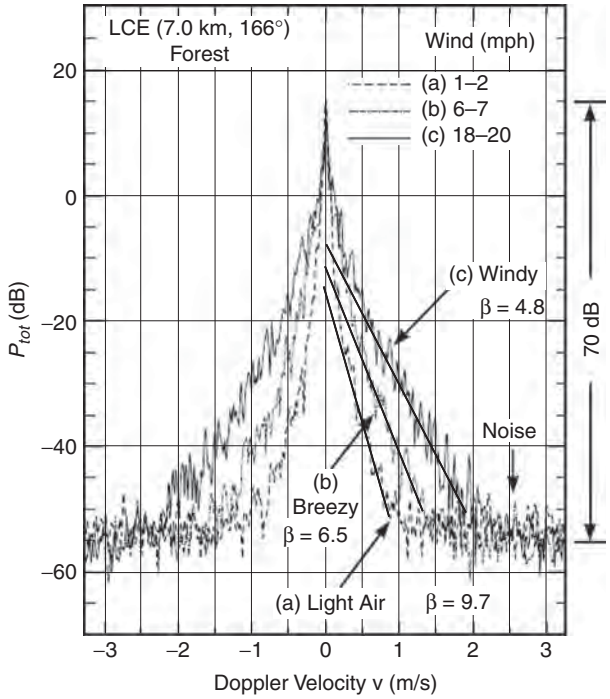


FIGURE 2.13 Measured spectra of clutter from forest. Several wind speeds and an estimated value of β have been added. (*after J.B. Billingsley*[9] *© William Andrew Publishing Inc. 2002*)

A comparison between the gaussian and the exponential models on a linear scale, as shown in Figure 2.14, indicates that the difference in spectral width at even very low levels (–80 dB) is no more than about a factor of 2. For many analyses, this would most likely be insignificant compared to the added clutter spectral spreading caused by scanning modulation. Thus, in many cases, the simple gaussian model can continue to be used in MTI and MTD performance analysis. In case of doubt, the spectral spread of the gaussian model could be doubled to assess the available margin.

Nathanson and Reilly[10] have shown that the clutter spectral width of rain is primarily due to a turbulence and wind shear (change in wind velocity with altitude). Measurements show a typical average value of σ_{vt} = 1.0 m/s for turbulence and σ_{vs} = 1.68 m/(s/km) for wind shear. A convenient equation is $\sigma_{vs} = 0.04 \cdot R \cdot \theta_{el}$ m/s for the effect of wind shear, provided the rain fills the vertical beam. Here R is the range to the weather, in nautical miles, and θ_{el} is the one-way half-power vertical beamwidth, in degrees. Thus, for example, σ_{vs} of rain viewed at 25 nmi with a vertical beamwidth of 4° would be σ_{vs} = 4.0 m/s. The total spectral spread is then $\sigma_v = \sqrt{\sigma_{vt}^2 + \sigma_{vs}^2} = \sqrt{1.0^2 + 4.0^2} = 4.1$ m/s. Rain and chaff also have an average velocity, in addition to the spectral spread noted above, which must be taken into account when designing an MTI system.

The clutter spectral width in meters per second is independent of the radar frequency. The standard deviation of the clutter power spectrum σ_f, in hertz, is

$$\sigma_f = \frac{2 \cdot \sigma_v}{\lambda} \quad \text{Hz} \tag{2.14}$$

where λ is the transmitted wavelength, in meters; and σ_v is the clutter standard deviation, in meters per second.

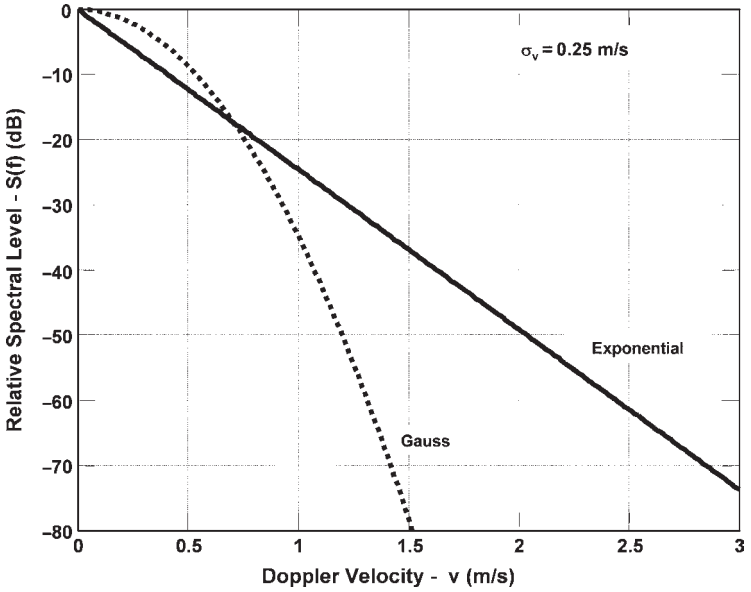


FIGURE 2.14 Comparison of gaussian and exponential spectra on linear velocity scale

Antenna scanning also causes a spread of the clutter power spectrum due to the amplitude modulation of the echo signals by the two-way antenna pattern.[11] The resulting clutter standard deviation is

$$\sigma_f = \frac{\sqrt{\ln 2}}{\pi} \cdot \frac{f_r}{n} = 0.265 \cdot \frac{f_r}{n} \quad \text{Hz} \tag{2.15}$$

where f_r is the PRF and n is the number of hits between the one-way 3-dB points of the antenna pattern. This equation was derived from a gaussian beam shape but is essentially independent of the actual beam shape or aperture illumination function used.

The clutter spectral spread due to scanning, normalized to the PRF, is

$$\sigma_f T = \frac{0.265}{n} \tag{2.16}$$

where $T = 1/\text{PRF}$ is the interpulse period.

The combined spectral effects of internal clutter motion and antenna scanning modulation must be obtained as the convolution of the individual spectra. When both spectra are gaussian in shape, the resulting spectrum remains gaussian with a standard deviation that is the square-root of the sum of the squares of the individual standard deviations.

By integrating the two-sided tails of the gaussian and exponential spectra, outside a multiple, *k,* of the standard deviation of the spectra, a rough, but conservative, estimate can be found of how wide the MTI notch must be to achieve a required improvement factor *I.* Such a curve is shown in Figure 2.15 based on the clutter spectra shown in Figure 2.14. Although this approach would only be strictly correct for an ideal MTI filter with a step-function passband, it can serve as a preliminary guideline for the MTI filter design.

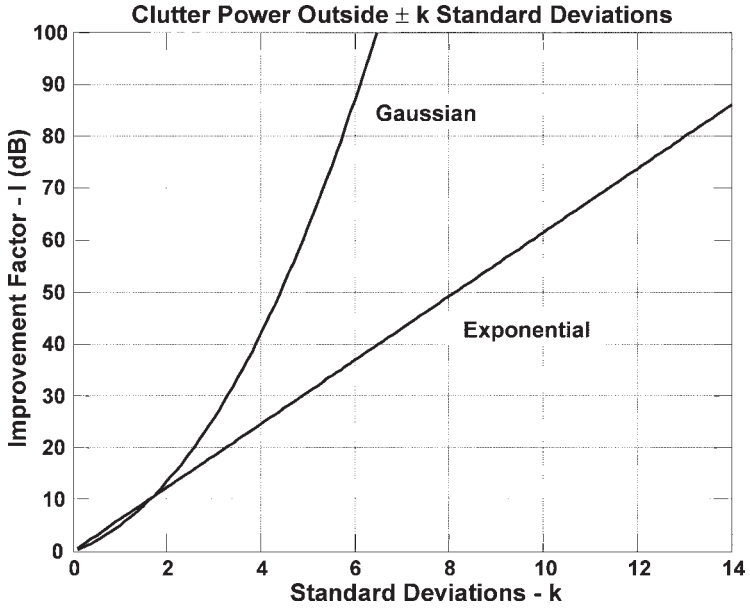


FIGURE 2.15 Clutter power in two-sided tails of spectrum vs. multiple of standard deviation

Amplitude Characteristics. To predict the performance of an MTI system, the power of the clutter returns with which a target must compete should be known. The amplitude of the clutter returns depends on the size of the resolution cell of the radar, the frequency of the radar, and the reflectivity of the clutter. The expected radar cross section of clutter can be expressed as the product of a reflectivity factor and the volume or area of the resolution cell.

For surface clutter, as viewed by a surface-based radar,

$$\bar{\sigma} = A_c \cdot \sigma^0 = R \cdot \theta_{az} \cdot \frac{c \cdot \tau}{2} \cdot \sigma^0 \tag{2.17}$$

where $\bar{\sigma}$ is the average radar cross section, in square meters; A_c is the area of clutter illuminated, in square meters; R is the range to clutter, in meters; θ_{az} is the one-way half-power azimuthal beamwidth, in radians; c is the speed of propagation, 300 million m/s; τ is the half-power radar pulse length (after the matched filter), in seconds; and σ^0 is the average clutter reflectivity factor, in square meters per square meter.

For volumetric clutter, such as chaff or rain, the average cross section is

$$\bar{\sigma} = V_c \cdot \eta = R \cdot \theta_{az} \cdot \theta_{el} \cdot H \cdot \frac{c \cdot \tau}{2} \cdot \eta \tag{2.18}$$

where V_c is the volume of clutter illuminated (m^3) and η is the clutter reflectivity factor (m^2/m^3). The volume V_c is computed from the height extent of clutter H (meters), the azimuth extent of the clutter $R \cdot \theta_{az}$, and the radar range resolution cell τ. If the clutter completely fills the vertical beam, then $H = R \cdot \theta_{el}$, where θ_{el} is the elevation beamwidth. R is the range to the clutter (meters) and c is the speed of propagation.

It should be noted that for land clutter σ^0 can vary considerably from one resolution cell to the next. A typical distribution of σ^0, taken from Barton,[12] is shown in Figure 2.16. Typical values for σ^0 and η from the same reference are given in Table 2.2. Additional results for clutter reflectivity are found in Billingsley.[9]

TABLE 2.2 Typical Values of Clutter Reflectivity*

Clutter	Reflectivity, λ, m η, $(m)^{-1}$	Conditions	Band λ, m	Clutter Parameters for Typical Conditions L 0.23	S 0.1	C 0.056	X 0.032
Land (excluding point clutter)	$\sigma^0 = \frac{0.00032}{\lambda}$ (worst 10 percent)	.....	σ^0 dB	–29	–25	–22	–20
Point clutter	$\sigma = 10^4$ m^2	.....	σ m^2	10^4	10^4	10^4	10^4
Sea (Beaufort scale K_B, angle E)	σ^0 dB = $-64 + 6K_B +$ (sin E)dB $- \lambda$ dB	Sea state 4 (6-ft waves, rough); $E = 1°$	σ^0 dB	–51.5	–47.5	–44.5	–42.5
Chaff (for fixed weight per unit volume)	$\eta = 3 \times 10^{-8}\ \lambda$	.....	η m^{-1}	7×10^{-9}	3×10^{-9}	1.7×10^{-9}	10^{-9}
Rain (for rate r mm/h)	$\eta = 6 \times 10^{-14}\ r^{1.6}\ \lambda^{-4}$ (matched polarization)	$r = 4$ mm/h	η m^{-1}	2×10^{-10}	5×10^{-9}	5×10^{-8}	5×10^{-7}

*From Barton[12]

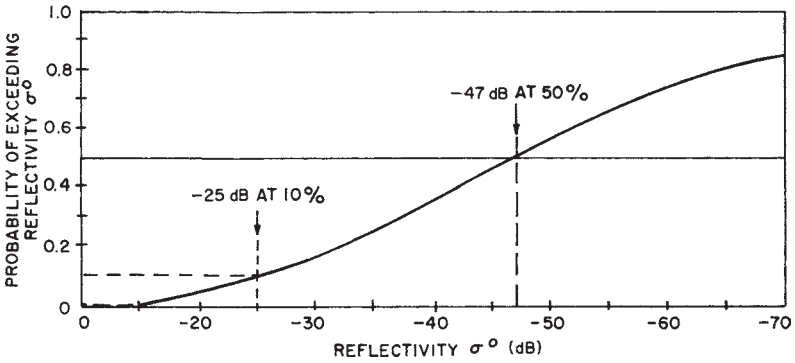


FIGURE 2.16 Distribution of reflectivity for ground clutter typical of heavy clutter at S band (*after D. K. Barton*[12] *© IEEE 1967*)

Because of the imprecision in predicting σ^0 and η, these equations do not include an antenna beam-shape factor. For the measurement of the reflectivity of rain, references on radar meteorology present more precise equations.[13]

In addition to distributed clutter targets, there are many targets that appear as *points,* such as radio towers, water tanks, and buildings. These point targets typically have a radar cross section of 10^3 to 10^4 m^2 with typical densities as shown later in Figure 2.18. This graph is from Billingsley[9] and the additional points indicated by an asterisk are from Ward.[14]

Figure 2.17*a* shows a PPI display of all clutter observed with a surveillance radar with a 1.3° by 2-μs resolution cell in the mountainous region of Lakehead, Ontario, Canada. (The PPI range is set for 30 nmi.) Clutter that exceeds the minimum-discernible signal (MDS) level of the radar by 60 dB is shown in Figure 2.17*b*.

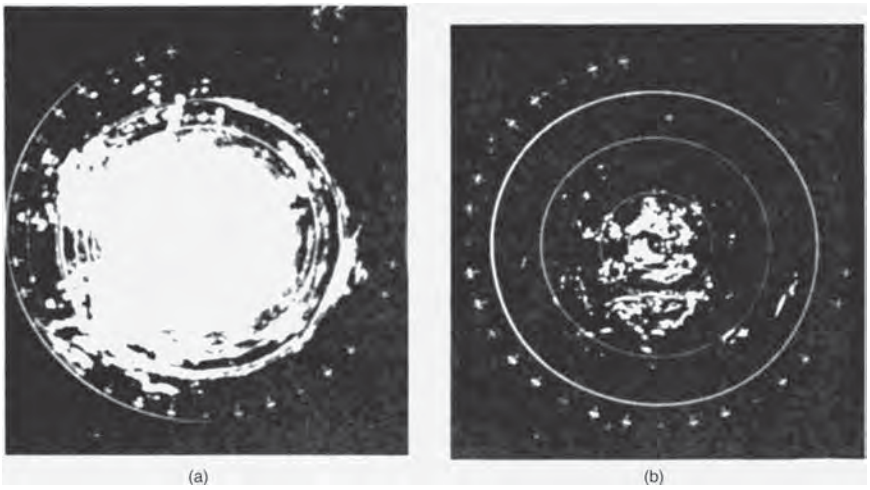


FIGURE 2.17 PPI display, 30-nmi range of (*a*) all clutter at a mountainous site and (*b*) clutter that exceeds the system noise level by 60 dB

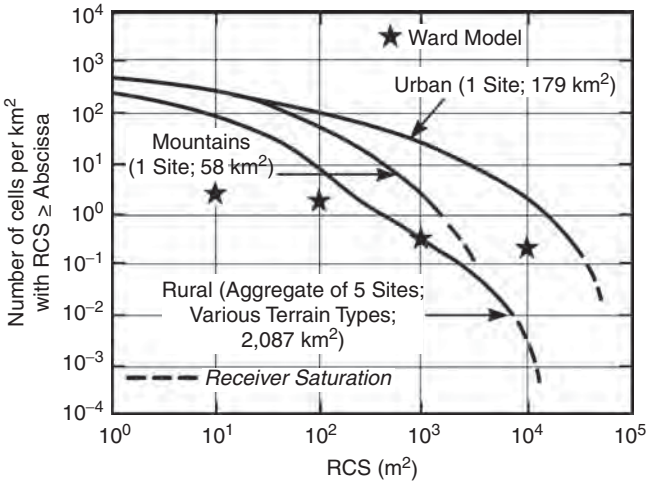


FIGURE 2.18 Typical densities of point clutter scatterers (*after J.B. Billingsley*[9] *© William Andrew Publishing Inc. 2002*)

Note that the clutter in Figure 2.17*b* is very spotty in character, including the strong fixed-point targets and returns from extended targets. It is significant that the extended targets are no longer very extended. The face of a mountain at 10 mi from 5 to 7 o'clock is only a line. If the MTI system were incapable of displaying an aircraft while it was over the mountain face, it would display the aircraft on the next scan of the antenna because the aircraft would have moved either farther or nearer. The PPI does not have a resolution that approaches the resolution of the signal processing circuits of this radar. Thus, the apparent extended clutter has many weak areas not visible in these photographs, where targets could be detected by virtue of an MTI radar's interclutter visibility (defined in Section 2.5).

2.5 DEFINITIONS

The IEEE Standard Radar Definitions[15] provide useful definitions for many of the quantities needed to quantify MTI and MTD performance, but in some cases, the vagueness of the original definition and the lack of distinction between performance against distributed clutter versus point clutter returns have led to ambiguous interpretations of several terms. In this section, the major definitions will be reviewed and annotated to attempt to clarify some of these potential ambiguities. For each term, the IEEE definition, when available, will be quoted along with a subsequent discussion.

Improvement Factor The IEEE definition of Improvement Factor reads:

moving-target-indication (MTI) improvement factor: The signal-to-clutter power ratio at the output of the clutter filter divided by the signal-to-clutter power ratio at the input to the clutter filter, averaged uniformly over all target radial velocities of interest. Synonym: **clutter improvement factor.**

This definition assumes that clutter is distributed homogeneously across many range cells. In this case, the above definition is equally valid before and after pulse compression. Against point clutter this definition only applies after pulse compression and may result in a different value of the improvement factor. The real difficulty with this definition is, however, the lack of a precise definition of the doppler velocity interval, which is to be used for the required "uniform" averaging. Originally, this averaging was assumed to involve multiple PRF intervals based on classical low PRF radars using a single MTI filter. It was for this reason that the MTI Improvement Factor definition (I) provided in the 2nd edition of this *Radar Handbook* used the noise gain of the doppler (MTI) filter as the normalizing factor. The increased use of pulse doppler filter banks in modern radar has, however, led to a use of the IEEE definition where the averaging of the signal-to-clutter ratio improvement is performed only across a narrow region around the peak of the doppler filter response. In this case, the coherent integration gain of the doppler filter is automatically added to the conventional MTI improvement factor value and much better radar performance is indicated.

Since a definition of clutter suppression is often needed, which quantifies the inherent radar stability limitations, apart from any additional coherent gain, it is sometimes preferable to use the IEEE definition of clutter attenuation. In this chapter, *improvement factor* and *clutter attenuation* will be used synonymously. When the coherent gain of the doppler filter is included, the term *signal-to-clutter ratio* improvement will be used.

Clutter Attenuation The IEEE definition reads

> **clutter attenuation (CA):** In moving-target indication (MTI) or doppler radar, the ratio of the clutter-to-noise ratio at the input to the processor, to the clutter-to-noise ratio at the output. *Note:* In MTI, a single value of **CA** will be obtained, while in doppler radar the value will generally vary over the different target doppler filters. In MTI, **CA** will be equal to MTI improvement factor if the targets are assumed uniformly distributed in velocity. *See also:* **MTI improvement factor.**

Here, it will be assumed that "processor" refers to the MTI filter or a single doppler filter in a pulse doppler filter bank. Based on this definition, the clutter attenuation is given by

$$CA = \frac{P_{\text{CIN}}}{P_{\text{COUT}}} \cdot \frac{P_{\text{NOUT}}}{P_{\text{NIN}}} \tag{2.19}$$

where P_{CIN} and P_{COUT} are the clutter power at the input and output of the MTI filter, respectively, and P_{NIN} and P_{NOUT} are the corresponding noise powers. As noted in the IEEE definition, the value of CA will most likely differ from filter to filter in a doppler filter bank due to specific clutter and filter response characteristics.

In the discussion above, the assumption was implicitly made that clutter returns are stationary and distributed in range. The above definitions will be equally valid before and after pulse compression. For a single piece of point clutter, as often used in actual radar stability measurements, the definition of clutter attenuation would have to be changed as follows to provide identical results:

> **clutter attenuation (CA), point clutter:** In moving-target indication (MTI) or Doppler radar, the ratio of the total energy in the received point clutter return at the input to the processor, to the total energy in the point clutter residue at the output of the processor, multiplied by the noise gain of processor.

The clutter attenuation against point clutter based on this definition will be the same before or after pulse compression and will also be identical to the value of CA obtained against distributed clutter with identical spectral characteristics.

For the practical measurement of CA against a single piece of point clutter (i.e., corner reflector), the total energy must be integrated, per the above definition, at the input and output of each doppler filter. The calculation of the energy is best performed prior to pulse compression since the precise duration of the uncompressed pulse, and therefore the integration window, is accurately known. If done after pulse compression, uncertainties in the integration of energy may arise due to the transient response of the pulse compression filter.

Signal-to-Clutter Ratio Improvement (I_{SCR}) For a system employing multiple doppler filters, such as the MTD, each doppler filter will also have a coherent gain, $G_C(f)$, which at the filter peak has a value $G_{C,\max}$. The coherent gain of a doppler filter is equal to the increase in signal-to-thermal-noise ratio between the input and the output of the filter due to the coherent summation of individual target returns. Again these coherent gain values would usually differ from filter to filter due to potentially different doppler filter characteristics. These coherent gain values will include the filter mismatch loss but not the straddling losses between adjacent filters. The product of the clutter attenuation, CA_i, and the coherent gain, $G_{C\max,i}$, for the i'th doppler filter becomes the definition of the signal-to-clutter ratio (SCR) improvement:

$$I_{\mathrm{SCR},i} = CA_i \cdot G_{C\max,i} \tag{2.20}$$

This quantity was not included in the *IEEE Dictionary,*[15] but the following definition is commonly used:

> **signal-to-clutter ratio improvement: (I_{SCR}):** The ratio of the signal-to-clutter ratio obtained at the output of the doppler filter bank to the signal-to-clutter ratio at the input to the filter bank computed as a function of target doppler frequency.

This definition does not include any doppler averaging across the individual filters, and the definition does not provide a single figure of merit for a radar doppler processor because each filter may have different values of clutter attenuation and coherent gain.

Since each doppler filter has a coherent gain that is a function of target doppler, an average value of signal-to-clutter improvement can be defined by averaging all filters over its respective range of target dopplers:

$$\bar{I}_{\mathrm{SCR}} = \frac{1}{f_N - f_0}\left[\begin{array}{l}\displaystyle\int_{f_0}^{f_1} CA_0 \cdot G_{C,0}(f)\cdot df + \int_{f_1}^{f_2} CA_1 \cdot G_{C,1}(f)\cdot df \\ \\ +\ldots\ldots\ldots\ldots + \displaystyle\int_{f_{N-1}}^{f_N} CA_{N-1}\cdot G_{C,N-1}(f)\cdot df\end{array}\right] \tag{2.21}$$

The specific frequencies could logically be chosen as the crossover between individual doppler filters. This calculation will now include the effect of a target doppler

straddling loss and would represent a single figure-of-merit for a doppler processor. To simplify this calculation the *average signal-to-clutter improvement* may be defined as the finite sum

$$\bar{I}_{\mathrm{SCR}} = \frac{1}{N}\sum_{i=0}^{N-1} CA_i \cdot G_{C\max,i} \tag{2.22}$$

to which the doppler straddling loss would have to be added.

Subclutter Visibility (SCV) The IEEE definition of subclutter visibility is

> **Subclutter visibility:** The ratio by which the target echo power may be weaker than coincident clutter echo power and still be detected with specified detection and false-alarm probabilities. *Note:* Target and clutter powers are measured on a single pulse return and all target velocities are assumed equally likely.

The subclutter visibility (SCV) of a radar system is a measure of its ability to detect moving-target signals superimposed on clutter signals. A radar with 20 dB SCV can detect an aircraft flying over clutter whose signal return is 100 times stronger. Note that it is implicitly assumed in the above definition that signal and clutter are both observed after pulse compression. The SCV of two radars cannot necessarily be used to compare their performance while operating in the same environment, because the target-to-clutter ratio seen by each radar is proportional to the size of the radar resolution cell and may be a function of frequency. Thus, a radar with a 10-μs pulse length and a 10° beamwidth would need 20 dB more subclutter visibility than a radar with a 1-μs pulse and a 1° beamwidth for equal performance in a distributed clutter environment.

The subclutter visibility of a radar, when expressed in decibels, is less than the improvement factor by the clutter visibility factor V_{oc} (see definition below).

Interclutter Visibility (ICV) The IEEE definition is

> **interclutter visibility:** The ability of a radar to detect moving targets that occur in resolution cells among patches of strong clutter; usually applied to moving target indication (MTI) or pulsed-Doppler radars. *Note:* The higher the radar range and/or angle resolution, the better the interclutter visibility.

The interclutter visibility (ICV) of a radar is a measure of its capability to detect targets between points of strong clutter by virtue of the ability of the radar to resolve the areas of strong and weak clutter. A radar with high resolution makes available regions between points of strong clutter where the target-to-clutter ratio will be sufficient for target detection even though the SCV of the radar (based on average clutter) may be relatively low. To achieve ICV, a mechanism must be furnished to provide CFAR operation against the residue from strong clutter. This CFAR is provided in older MTI system by IF limiting and, in the MTD implementation, through the use of high-resolution clutter maps. A quantitative definition of interclutter visibility has not yet been formulated.

Filter Mismatch Loss The IEEE definition is

> **filter mismatch loss:** The loss in output signal-to-noise ratio of a filter relative to the signal-to-noise ratio from a matched filter.

The maximum signal-to-noise ratio available from an N-pulse filter is N times the signal-to-noise ratio of a single pulse, assuming all pulses have equal amplitude. When weighting is applied to reject clutter and control the filter sidelobes, the peak output signal-to-noise ratio is reduced. The filter mismatch loss is the amount by which the peak-output signal-to-noise ratio is reduced by the use of weighting. A three-pulse MTI filter using binomial weights has a filter mismatch loss of 0.51 dB. The mismatch loss for the binomial-weighted four-pulse canceler is 0.97 dB.

Clutter Visibility Factor (V_{oc}) The IEEE definition is

> **clutter detectability factor:** The predetection signal-to-clutter ratio that provides stated probability of detection for a given false alarm probability in an automatic detection circuit. *Note:* In MTI systems, it is the ratio after cancellation or doppler filtering.

The clutter visibility factor is the ratio by which the target signal must exceed the clutter residue so that target detection can occur without having the clutter residue result in false-target detections. The system must provide a threshold that the targets will cross and the clutter residue will not cross.

2.6 IMPROVEMENT FACTOR CALCULATIONS

Using Barton's approach,[17] the maximum improvement factor I against zero-mean clutter with a gaussian-shaped spectrum for different implementations of the finite-impulse-response binomial-weight MTI canceler (see Section 2.8) is

$$I_1 \approx 2\left(\frac{f_r}{2\pi\sigma_f}\right)^2 \tag{2.23}$$

$$I_2 \approx 2\left(\frac{f_r}{2\pi\sigma_f}\right)^4 \tag{2.24}$$

$$I_3 \approx \frac{4}{3}\left(\frac{f_r}{2\pi\sigma_f}\right)^6 \tag{2.25}$$

where I_1 is the MTI improvement factor for the single-delay coherent canceler; I_2 is the MTI improvement factor for the dual-delay coherent canceler; I_3 is the MTI improvement factor for the triple-delay coherent canceler; σ_f is the rms frequency spread of the gaussian clutter power spectrum, in hertz; and f_r is the radar repetition frequency, in hertz. When the values of σ_f for scanning modulation in Eq. 2.15 are substituted in the above equations for I, the limitation on I due to scanning is

$$I_1 \approx \frac{n^2}{1.39} \tag{2.26}$$

$$I_2 \approx \frac{n^4}{3.84} \tag{2.27}$$

$$I_3 \approx \frac{n^6}{16.0} \tag{2.28}$$

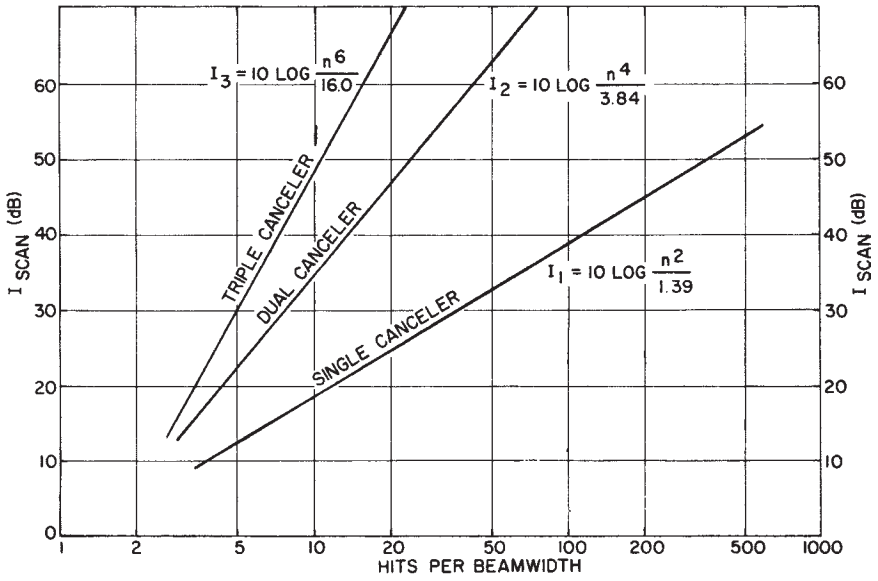


FIGURE 2.19 Theoretical MTI improvement factor due to scan modulation: gaussian antenna pattern: n = number of pulses within the one-way half-power beamwidth

These relationships are shown graphically in Figure 2.19. This derivation assumes a linear system. That is, it is assumed that the voltage envelope of the echo signals, as the antenna scans past a point target, is identical to the two-way antenna voltage pattern. This assumption of a linear system may be unrealistic for some practical MTI systems with relatively few hits per beamwidth, however, as discussed in Section 2.11.

The scanning limitation does not apply to a system that can step-scan, such as a phased array. Note, however, that sufficient pulses must be transmitted to initialize the filter before useful outputs may be obtained. For example, with a three-pulse binomial-weight canceler, the first two transmitted pulses initialize the canceler, and a useful output is not available until after the third pulse has been transmitted. Feedback or infinite impulse response (IIR) filters would not be used with a step-scan system because of the long transient settling time of the filters.

The limitation on I due to internal-clutter fluctuations can be determined by substituting the appropriate value of σ_f into Eqs. 2.23 to 2.25. By letting $\sigma_f = 2\sigma_v/\lambda$, where σ_v is the rms velocity spread of the clutter, the limitation on I can be plotted for different types of clutter as a function of the wavelength λ and the pulse repetition frequency f_r. This is done for one-, two-, and three-delay binomial-weight cancelers in Figure 2.20, Figure 2.21, and Figure 2.22. The values of V_B given are the first blind speed of the radar (or where the first blind speed V_B would be for a staggered PRF system if staggering were not used). The improvement factor shown in these figures for rain and chaff is based on the assumption that the average velocity of the rain and chaff has been compensated for so that the returns are centered in the canceler rejection notch. Unless such compensation is provided, the MTI offers little or no improvement for rain and chaff.

Two further limitations on I are the effect of pulse-to-pulse repetition-period staggering combined with clutter spectral spread from scanning and internal-clutter motion.

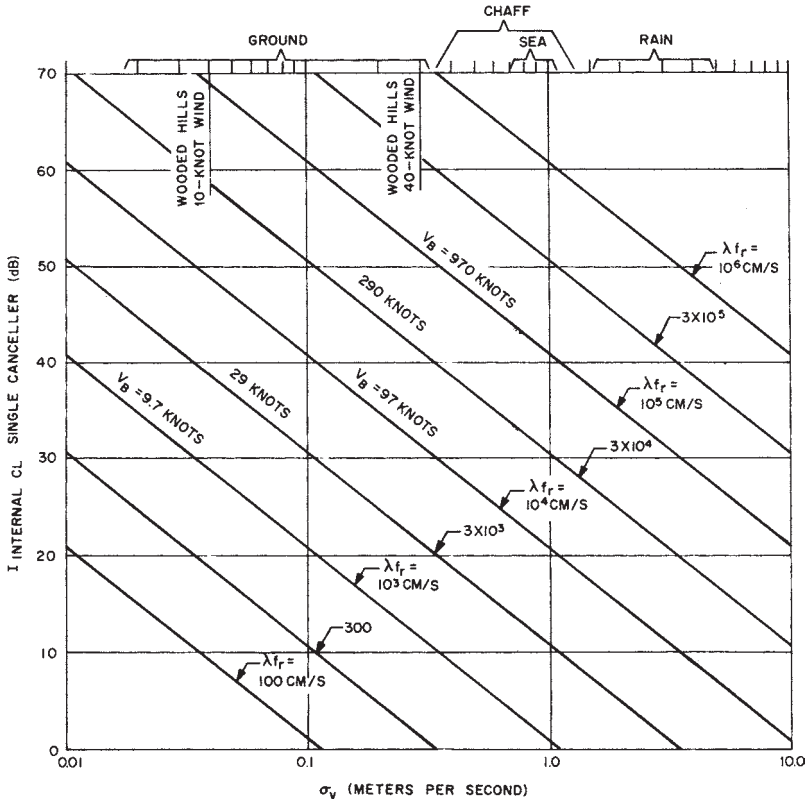


FIGURE 2.20 MTI improvement factor as a function of the rms velocity spread of clutter for a two-pulse binomial-weight canceler

These limitations, plotted in Figure 2.23 and Figure 2.24, apply to all cancelers, whether single or multiple. (The derivation of these limitations and a means of avoiding them by the use of time-varying weights are given in "Stagger Design Procedures" in Section 2.8.)

2.7 OPTIMUM DESIGN OF CLUTTER FILTERS

The statistical theory of detection of signals in gaussian noise provides the required framework for the optimum design of radar clutter filters. Such theoretical results are important to the designer of a practical MTI or MTD system, in that they establish upper bounds on the achievable performance in a precisely specified clutter environment. It should be noted, however, that owing to the extreme variability of the characteristics of real clutter returns (power level, doppler shift, spectrum shape, spectral width, etc.) any attempt to actually approximate the performance of such optimum filters for the detection of targets in clutter requires the use of adaptive methods. The adaptive methods must estimate the unknown clutter statistics and

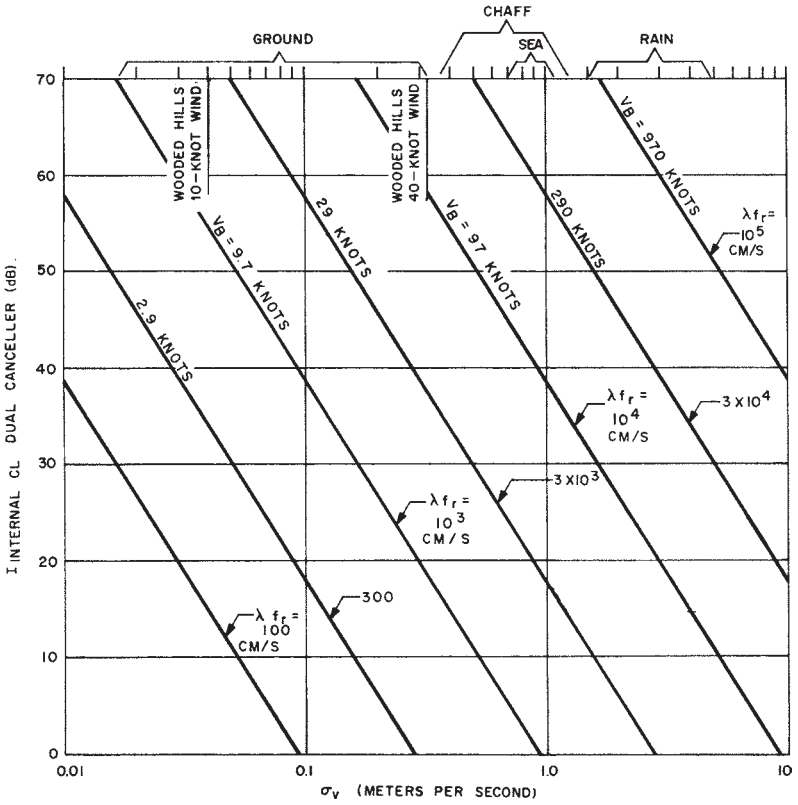


FIGURE 2.21 MTI improvement factor as a function of the rms velocity spread of clutter for a three-pulse binomial-weight canceler

subsequently implement the corresponding optimum filter. An example of such an adaptive MTI system is discussed in Section 2.14.

For a single radar pulse with a duration of a few microseconds, the doppler shift due to aircraft target motion is a small fraction of the signal bandwidth, and conventional MTI and pulse doppler processing are not applicable. It is well known that the classical single-pulse "matched" filter provides optimum radar detection performance when used in a white-noise background. Against clutter returns that have the same spectrum as the transmitted radar pulse, the matched filter is no longer optimum, but the potential improvement in the output signal-to-clutter ratio by designing a modified optimized filter is usually insignificant.

When the duration of the transmitted radar signal, whether CW or a repetitive train of N identical pulses, is comparable with or greater than the reciprocal of anticipated target doppler shifts, the difference between a conventional white-noise matched filter (or coherent integrator) and a filter optimized to reject the accompanying clutter becomes significant. The characteristics of the clutter are characterized by the covariance matrix Φ_C of the N clutter returns. If the power spectrum of the clutter is denoted

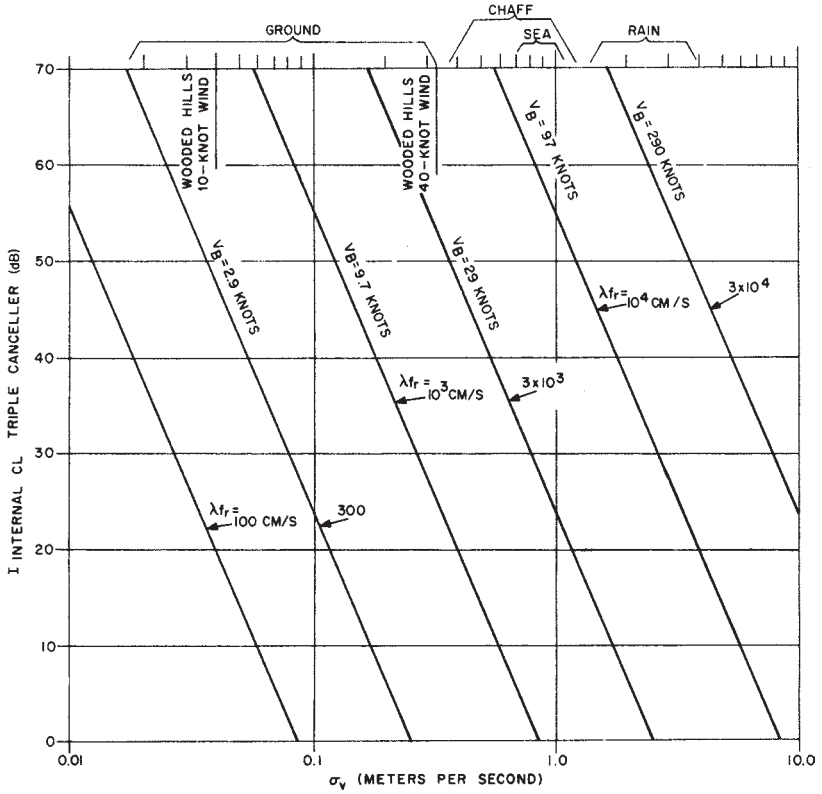


FIGURE 2.22 MTI improvement factor as a function of the rms velocity spread of clutter for a four-pulse binomial-weight canceler

$S_C(f)$ and the corresponding autocorrelation function is $R_C(t_i - t_j)$, then the elements of Φ_C are given by

$$\Phi_{ij} = R_C(t_i - t_j) \tag{2.29}$$

where t_i is the transmission time of the ith pulse. For example, for a gaussian-shaped clutter spectrum we have

$$S_C(f) = P_C \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_f} \cdot \exp\left[-\frac{(f - f_d)^2}{2 \cdot \sigma_f^2}\right] \tag{2.30}$$

where P_C is the total clutter power, σ_f is the standard deviation of the clutter spectral width, and f_d is the average doppler shift of the clutter. The corresponding autocorrelation function is

$$R_C(\tau) = P_C \ \exp\ (-4\pi\sigma_f^2\tau^2)\ \exp\ (-j2\pi f_d\tau) \tag{2.31}$$

where τ is the separation in time of two consecutive clutter returns.

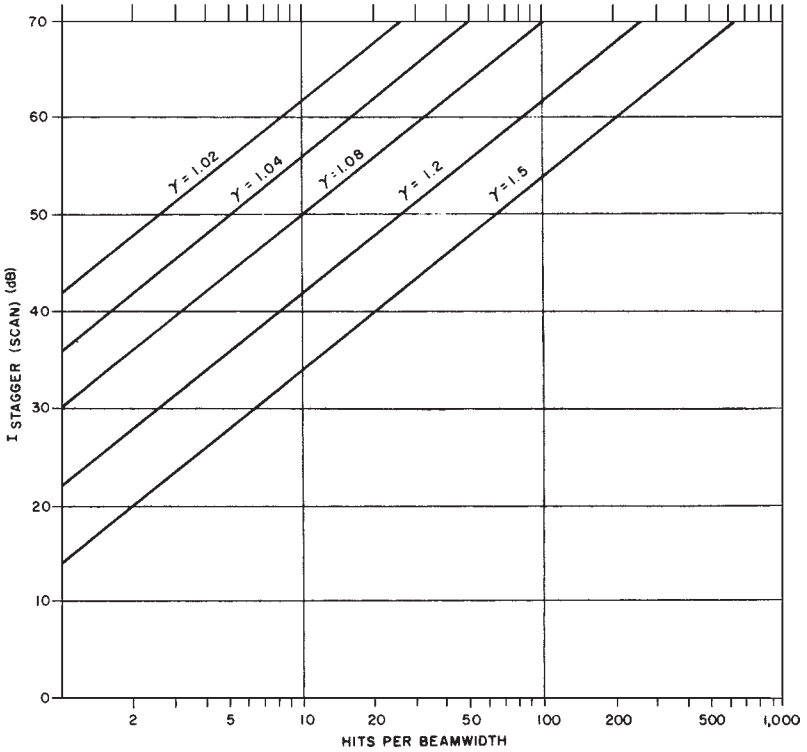


FIGURE 2.23 Approximate MTI improvement factor limitation due to pulse-to-pulse repetition-period staggering and scanning (all canceler figurations): I(dB) = 20 log [2.5n/ (γ − 1)]; γ = maximum period/ minimum period

For two pulses separated in time by the interpulse period T, the complex correlation coefficient between two clutter returns is

$$\rho_T = \exp\left(-4\pi\sigma_f^2 T^2\right) \cdot \exp\left(-j2\pi f_d T\right) \tag{2.32}$$

The second factor in this expression represents the phase shift caused by the doppler shift of the clutter returns.

For a known target doppler shift, the received target return can be represented by an N-dimensional vector:

$$\mathbf{s} = A_S \cdot \mathbf{f} \tag{2.33}$$

where A_S is the signal amplitude and the elements of the vector $\mathbf{f}$ are $f_i = \exp\,[j2\pi f_s t_i]$. On the basis of this description of signal and clutter, it has been shown[18] that the optimum doppler filter will have weights given by

$$\mathbf{w}_{\text{OPT}} = \Phi_{\mathbf{C}}^{-1} \cdot \mathbf{s} \tag{2.34}$$

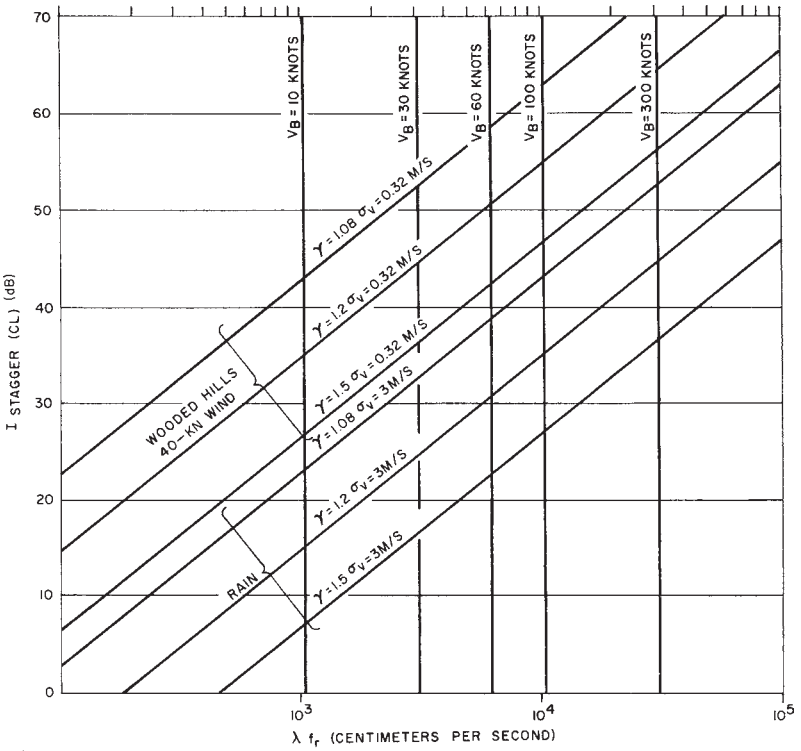


FIGURE 2.24 Approximate MTI improvement factor limitation due to pulse-to-pulse staggering and internal-clutter motion (all canceler configurations): I(dB) = 20 log [0.33/(λ − 1) ($\gamma f_r/\sigma_v$)]; γ = maximum period/minimum period

and the corresponding signal-to-clutter improvement is

$$I_{\text{SCR}} = \frac{\mathbf{w}_{\text{opt}}^T \mathbf{s} \cdot \mathbf{s}^{T^*} \mathbf{w}_{\text{opt}}^*}{\mathbf{w}_{\text{opt}}^T \mathbf{\Phi}_C \mathbf{w}_{\text{opt}}^*} \tag{2.35}$$

where the asterisk denotes complex conjugation and superscript T is the transposition operator. An example where the optimum performance is determined for the case of clutter at zero doppler having a gaussian-shaped spectrum with a normalized width of $\sigma_f T = 0.1$ is shown in Figure 2.25. In this case, a coherent processing interval of CPI = nine pulses was assumed, and the limitation due to thermal noise was ignored by setting the clutter level at 100 dB above noise.

It should be kept in mind that Eq. 2.34 for the optimum weights will yield a different result for each different target doppler shift, so that a large number of parallel filters would be needed to approximate the optimum performance even when the clutter characteristics are known exactly. As an example, the response of the optimum filter designed for one particular target doppler frequency labeled as point A in Figure 2.25 is shown in a broken line. At approximately ±5% from the design doppler, the performance starts to fall significantly below the optimum.

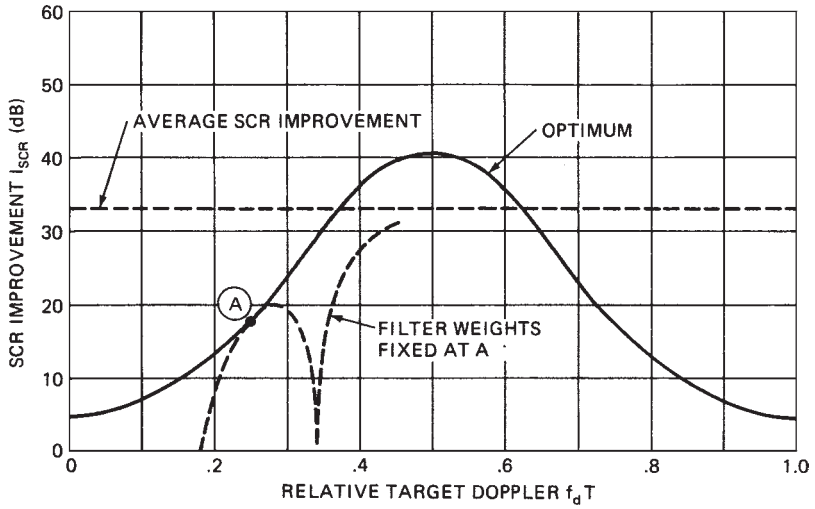


FIGURE 2.25 Optimum signal-to-clutter ratio improvement I_{SCR}) for gaussian-shaped clutter spectrum and a CPI of nine pulses; clutter-to-noise ratio, 100 dB

Also shown in Figure 2.25 is a horizontal line labeled "average SCR improvement." This indicates the level corresponding to the average of the optimum SCR curve across one doppler interval and may be considered as a figure of merit for a multiple-filter doppler processor, somewhat analogous to the MTI improvement factor defined for a single doppler filter. In Figure 2.26, the optimum average I_{SCR} has been computed for several different values of the CPI as a function of the normalized spectrum width. These results may be used as a point of reference for practical doppler

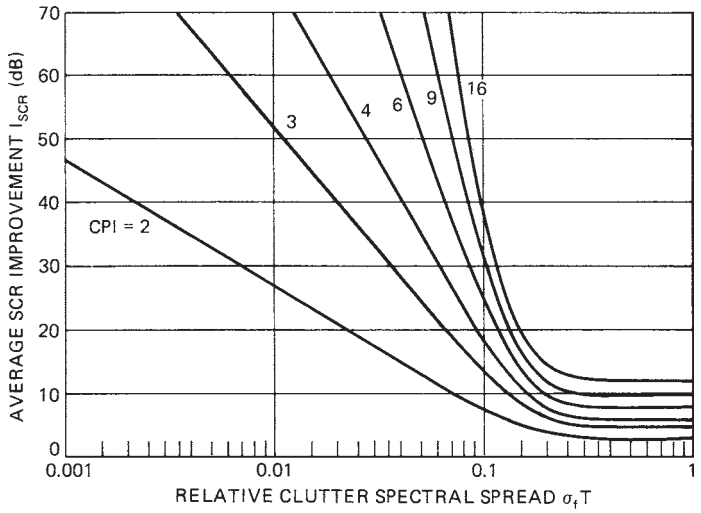


FIGURE 2.26 Reference curve of optimum average SCR improvement for a gaussian-shaped clutter spectrum

processor designs as discussed in Section 2.9. Note that for $\sigma_f T \approx 1$ the average SCR improvement is due only to the coherent integration of all the pulses in the CPI.

An MTI filter can also be designed based on the criterion of maximizing the signal-to-clutter improvement at a specific target doppler. However, such a design will usually provide suboptimum performance at all other target dopplers. The single exception is the two-pulse MTI canceler,[19] which provides optimum performance for all target dopplers.

A more attractive approach for designing an optimum MTI filter is to maximize its improvement factor (or clutter attenuation). To design an optimum MTI filter using improvement factor as the criterion, the covariance matrix of the clutter returns, as given by Eq. 2.29, is again the starting point. As shown by Capon,[20] the weights of the optimum MTI filter are found as the eigenvector corresponding to the smallest eigenvalue of the clutter covariance matrix and the MTI improvement factor is equal to the inverse of the smallest eigenvalue. The optimum improvement factor for the three models for the spectrum of land clutter introduced in Section 2.4 have been computed based on this above approach.

For the gaussian clutter spectrum, the optimum improvement factor is shown in Figure 2.27 as a function of the rms relative spectrum width, assuming zero mean for the spectrum. Calculations are shown for MTI cancelers of order $N = 2$ through 32.

For the polynomial clutter spectrum, the optimum improvement factor is shown in Figure 2.28, again as a function of the RMS relative spectrum width assuming zero mean for the spectrum.

Finally, for the exponential clutter spectrum model, the optimum improvement factor is shown in Figure 2.29, again as a function of the RMS relative spectrum width, assuming zero mean for the spectrum.

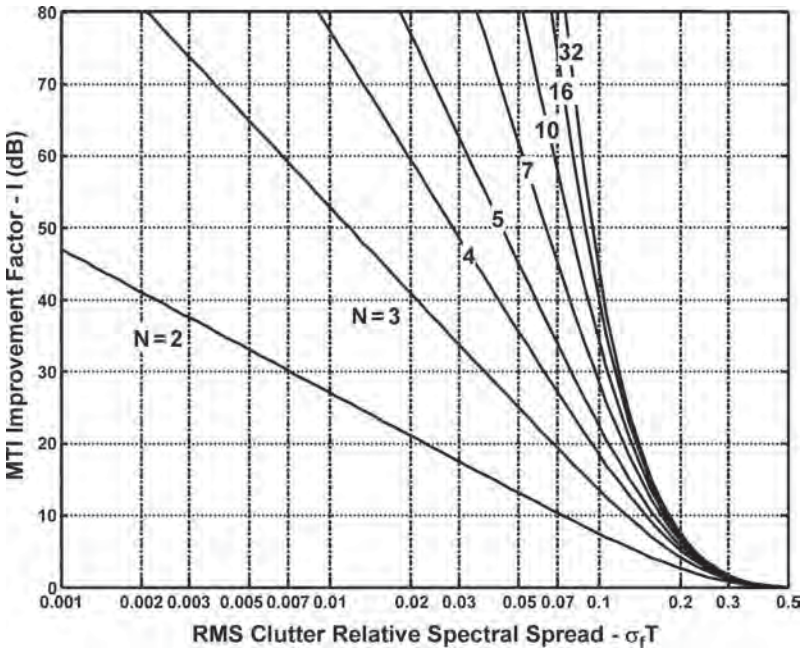


FIGURE 2.27 Optimum improvement factor for gaussian spectrum model

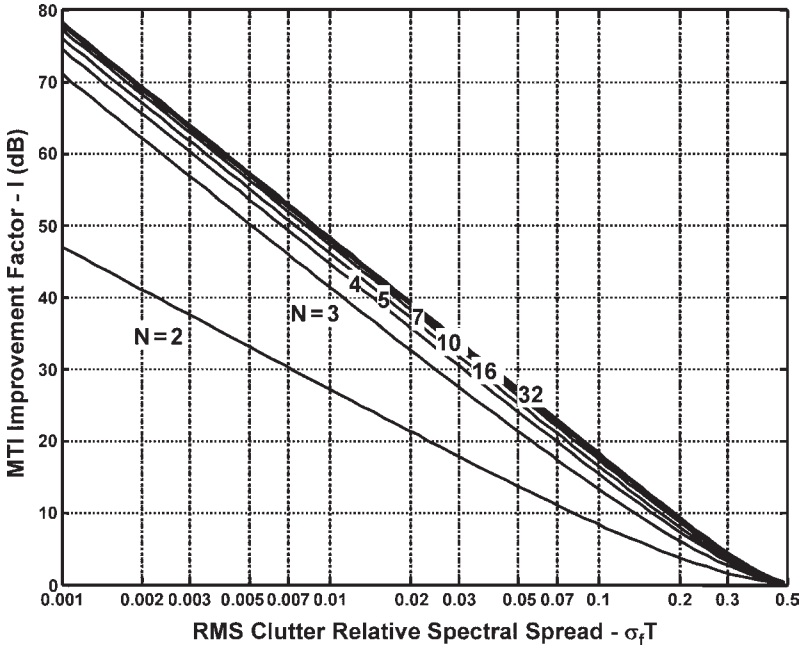


FIGURE 2.28 Optimum improvement factor for polynomial clutter spectrum model

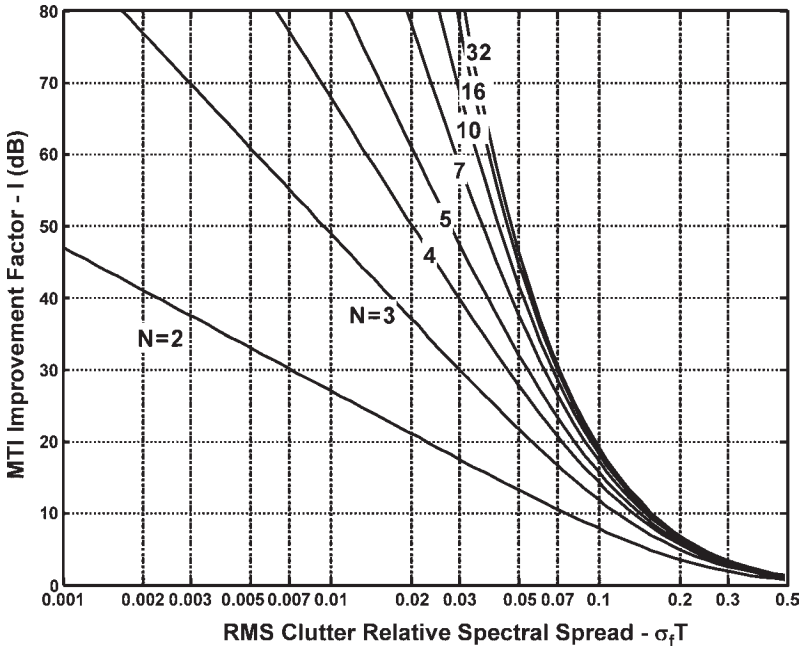


FIGURE 2.29 Optimum improvement factor for Billingsley's exponential spectrum model

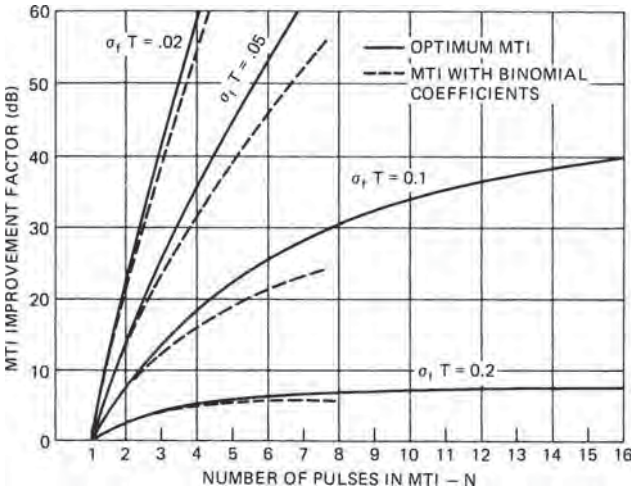


FIGURE 2.30 Comparison of MTI improvement factor of binomial-weight MTI and optimum MTI against a gaussian-shaped clutter spectrum

In Figure 2.30, the improvement factor of an MTI using the optimum weights is compared with the binomial coefficient MTI for different values of the relative clutter spectral spread and shown as a function of the number of pulses in the CPI. These results again assume a gaussian-shaped clutter spectrum. For typical numbers of pulses in the MTI (three to five), the binomial coefficients are remarkably robust and provide a performance which is within a few decibels of the optimum. Again, it should be noted that any attempt to implement an MTI canceler, which performs close to the optimum, would require the use of adaptive techniques that estimate the clutter characteristics in real time. If the estimate is in error, the actual performance may fall below that of the binomial-weight MTI canceler.

2.8 MTI CLUTTER FILTER DESIGN

The MTI block diagrams introduced by Figures 2.2 and 2.3 and whose response was discussed in detail in Section 2.3, considered a single-delay* canceler. It is possible to utilize more than one delay and to introduce feedback and/or feedforward paths around the delays to change the MTI system response to targets of different velocities. Filters with only feedforward paths are called finite impulse response (FIR) filters, and filters that incorporate feedback are called infinite impulse response (IIR) filters, or recursive filters. Multiple-delay cancelers have wider clutter rejection notches than single-delay cancelers. The wider rejection notch encompasses more of the clutter spectrum and thus increases the MTI improvement factor attainable with a given clutter spectral distribution.

* Delay is used here to represent an interpulse memory for an MTI filter. An FIR filter with one delay is a two-pulse filter. For feedback (IIR) filters, it is inappropriate to call them two-pulse (or three-pulse, etc.) filters because they require a number of pulses to reach steady-state.

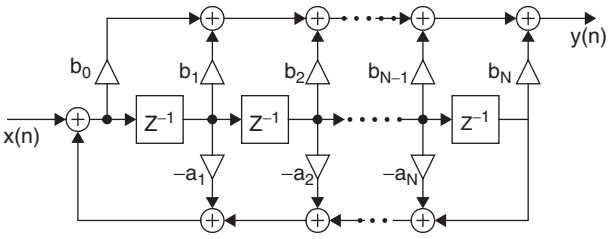


FIGURE 2.31 Direct Form 2 or canonical form of any MTI filter design

A general block diagram model applicable to any MTI filter is shown in Figure 2.31. This model has been denoted the "Direct Form 2," or the canonical form, in the terminology survey presented in Rabiner et al.[21]

It can be shown that an MTI filter as shown in Figure 2.31 can be divided into a cascade of second order sections as shown in Figure 2.32.

When a number of single-delay feedforward cancelers are cascaded in series, the overall filter voltage response is $k2^n \sin^n(\pi f_d T)$, where k is the target amplitude, n is the number of delays, f_d is the doppler frequency, and T is the interpulse period.[22] The cascaded single-delay cancelers can be rearranged as a transversal filter, and the weights for each pulse are the binomial coefficients with alternating sign: 1, −1 for two pulses; 1, −2, 1 for three pulses; 1, −3, 3, −1 for four pulses, and so on. Changes of the binomial feedforward coefficients and/or the addition of feedback modify the

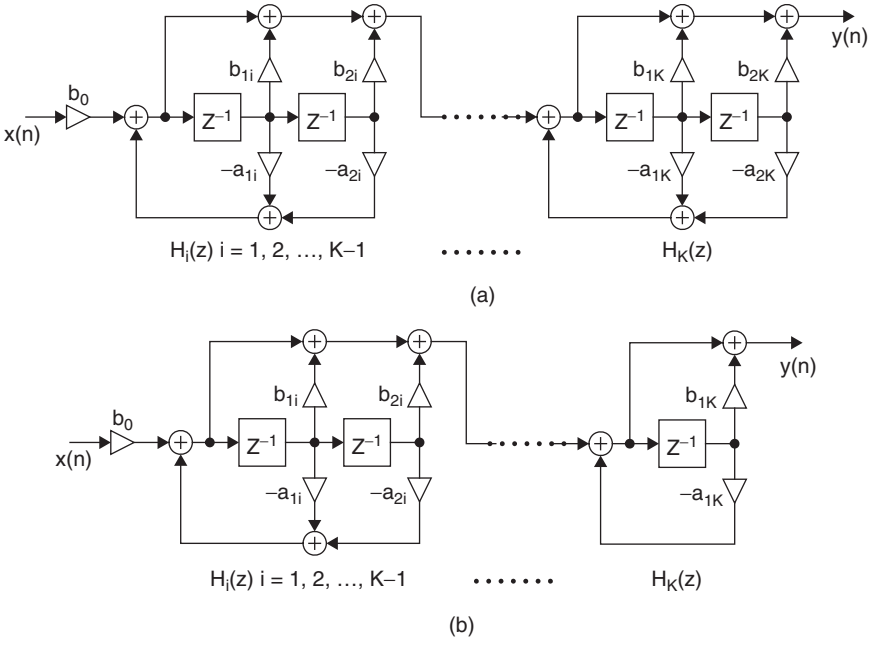


FIGURE 2.32 MTI shown as cascaded form of second order section: (*a*) is for even order and (*b*) is for odd order with first order section at end.

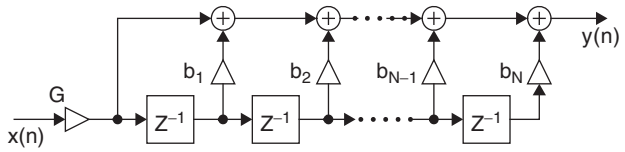


FIGURE 2.33 *N*th order FIR MTI canceler block diagram

filter characteristics. Within this chapter, reference to *binomial-weight cancelers* refers to cancelers with the $2^n \sin^n(\pi f_d T)$ transfer function. The block diagram of this type of MTI canceler is shown in Figure 2.33.

Figure 2.34 to Figure 2.36 represent typical velocity response curves obtainable from one-, two-, and three-delay cancelers. Shown also are the canceler configurations assumed, with corresponding *Z*-plane pole-zero diagrams. The *Z* plane is the comb-filter equivalent of the *S* plane[23] with the left-hand side of the *S* plane transformed to the inside of the unit circle centered at $Z = 0$. Zero frequency is at $Z = 1 + j0$. The stability requirement is that the poles of the *Z* transfer function lie within the unit circle. Zeros may be anywhere.

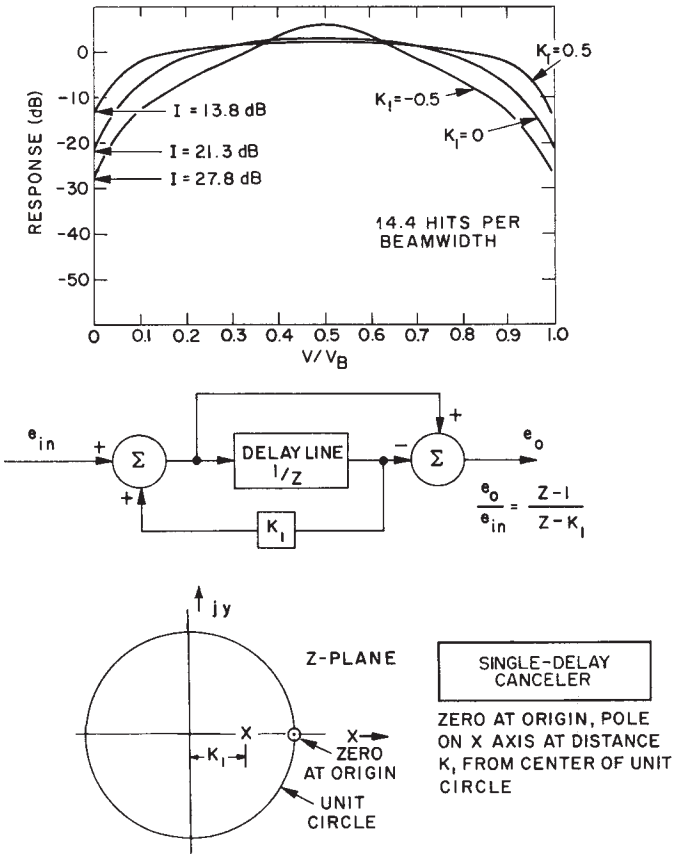


FIGURE 2.34 One-delay canceler

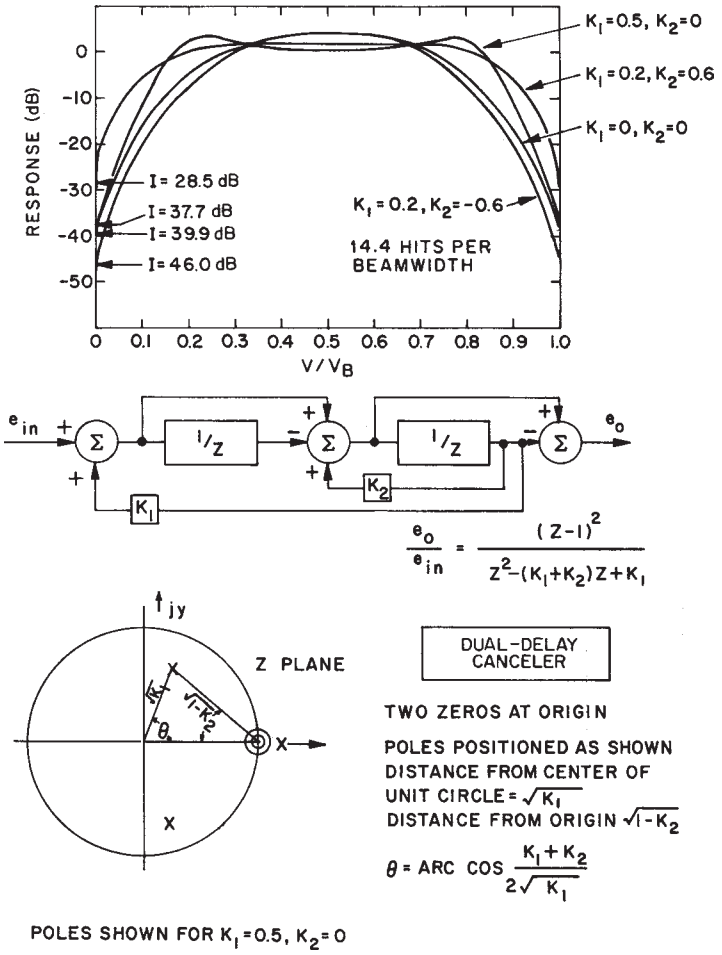


FIGURE 2.35 Two-delay canceler

These velocity response curves are calculated for a scanning radar system with 14.4 hits per one-way 3-dB beamwidth. An antenna beam shape of (sin *U*)/*U*, terminated at the first nulls, was assumed. The shape of these curves, except very near the blind speeds, is essentially independent of the number of hits per beamwidth or the assumed beam shape.

The ordinate labeled "response" represents the single-pulse signal-to-noise output of the MTI receiver relative to the signal-to-noise response of a normal linear receiver for the same target. Thus, all the response curves are normalized with respect to the noise power gain for the given canceler configuration. The intersection at the ordinate represents the negative decibel value of *I*, the MTI improvement factor for a point clutter target processed in a linear system.

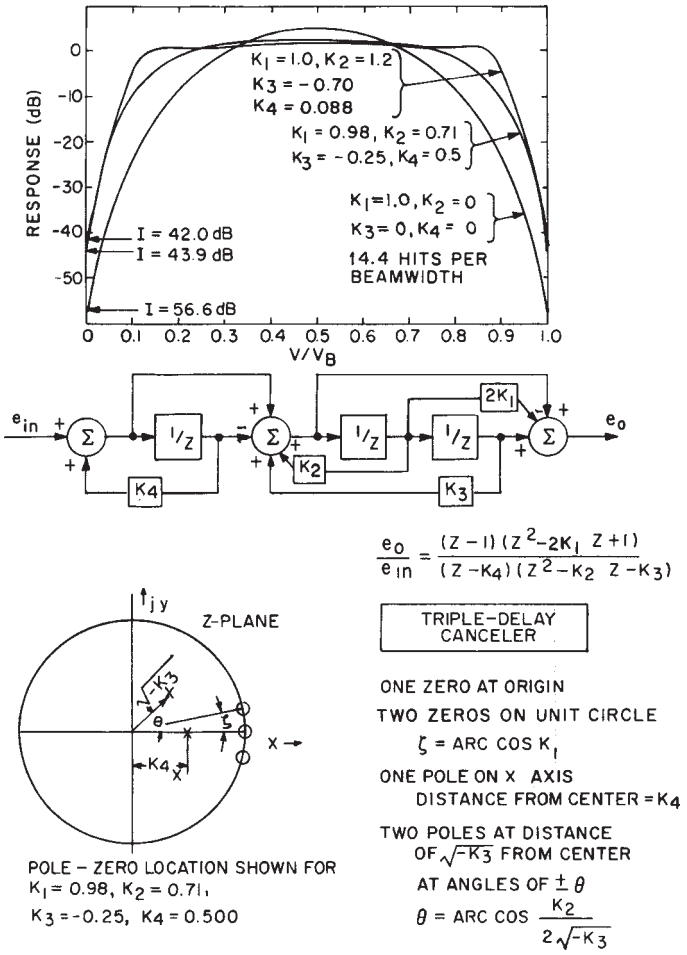


FIGURE 2.36 Three-delay canceler

Because these curves show the signal-to-noise response for each output pulse from the MTI canceler, the inherent loss incurred in a scanning radar with MTI processing due to the reduction of the effective number of independent pulses integrated[24] is not apparent. This loss is 1.4 dB for a 3-pulse canceler and 2.1 dB for a 5-pulse canceler, assuming a large number of pulses. If quadrature MTI channels (see Section 2.13) are not employed, there is an additional loss of 1½ to 3 dB.

The abscissa of these curves, V/V_B, represents the ratio of target velocity V to the blind speed $V_B = \lambda f_r/2$, where λ is the radar wavelength and f_r is the average PRF of the radar. The abscissa can also be interpreted as the ratio of the target doppler frequency to the average PRF of the radar.

The canceler configurations shown are not the most general feedforward, feedback networks possible. Pairs of delays are required to locate zeros and poles elsewhere

than on the real axis of the *Z*-plane. In the configurations shown, the zeros are constrained to the unit circle. To move the zeros off of the unit circle, which may be done to control the flatness of the filter passband response, requires a configuration similar to the elliptic filter configuration shown in Figure 2.46 later in this chapter. The triple-canceler configuration shown is such that two of the zeros can be moved around the unit circle in the *Z* plane. Moving the zeros can provide a 4 or 5 dB increase in the MTI improvement factor for specific clutter spectral spreads, as compared with keeping all three zeros at the origin.[25]

Note the width of the rejection notches for the different binomial-weight canceler configurations. If the −6 dB response relative to average response is used as the measuring point, the rejection is 24% of all target dopplers for the single canceler, 36% for the dual canceler, and 45% for the triple canceler. Consider the dual canceler: Eliminating 36% of the dopplers means limiting the system to a long-term average of 64% single-scan probability of detection. Feedback can be used to narrow the rejection notch without much degradation of *I*. If feedback is used to increase the improvement factor, the single-scan probability of detection becomes worse.

Figure 2.37 shows the improvement factor limitation due to scanning for cancelers with feedback. These curves were calculated assuming a (sin *U*)/*U* antenna pattern terminated at the first nulls.

The no-feedback curves shown in Figure 2.37 are almost indistinguishable from the theoretical curves derived for a gaussian pattern shown in Figure 2.19. (One of the curves showing the effect of feedback on the triple canceler is not straight because two of the three zeros are not at the origin but have been moved along the unit circle the optimum amount for 14 hits per beamwidth. Thus, at 40 hits per beamwidth, these two zeros are too far removed from the origin to be very effective.)

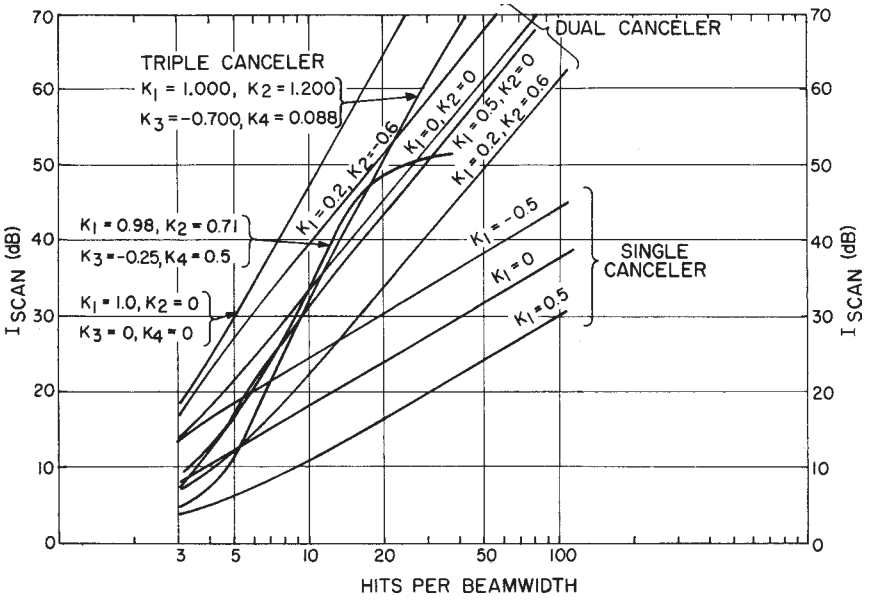


FIGURE 2.37 Improvement factor limitation due to scanning for cancelers with feedback

In theory, it is possible to synthesize almost any velocity response curve with digital filters.[26] As mentioned earlier, for each pair of poles and pair of zeros on the Z plane, two delay sections are required. The zeros are controlled by the feedforward paths and the poles by the feedback paths.

Velocity response shaping can be accomplished by the use of feedforward only without the use of feedback. The principal advantage of not using feedback is the excellent transient response of the canceler, an important consideration in a phased array or when pulse interference noise is present. If a phased array radar should use a feedback canceler, many pulses would have to be gated out after the beam has been repositioned before the canceler transient response has settled to a tolerable level. An initialization technique has been proposed[27] to alleviate this problem, but it provides only partial reduction in the transient settling time. If feedforward only is used, only three or four pulses have to be gated out after moving the beam. The disadvantage of using feedforward for velocity response shaping is that an additional delay, and therefore an additional transmit pulse, must be provided for each zero used to shape the response. Figure 2.38 shows the velocity response and Z-plane diagram of a feedforward-only, shaped-response, four-pulse canceler. Also shown are the velocity responses of a five-pulse feedforward canceler and a three-pulse feedback canceler. For the cancelers shown, the improvement factor capability of the three-pulse canceler is about 4 dB better than the shaped-response four-pulse feedforward canceler, independent of clutter spectral spread.

The five-pulse canceler response shown is a linear-phase[28] MTI filter described by Zverev.[29] The four zeros are located on the Z-plane real axis at +1.0, +1.0, −0.3575, and −2.7972. Much of the literature on filter synthesis describes linear-phase filters, but for MTI applications linear phase is of no importance. Almost identical filter responses can be obtained with nonlinear-phase filters that require fewer pulses, as shown in Figure 2.38. Because only a fixed number of pulses is available during the time on target, none should be wasted. Thus, one should choose the nonlinear-phase filter that uses fewer pulses.

Stagger Design Procedures. The interval between radar pulses may be changed to modify the target velocities to which the MTI system is blind. The interval may be changed on a pulse-to-pulse, dwell-to-dwell (each dwell being a fraction of the beamwidth), or scan-to-scan basis. Each approach has advantages. The advantages of the scan-to-scan method are that it is easier to build a stable transmitter, and multiple-time-around clutter is canceled in a power amplifier MTI system. The transmitter stabilization necessary for good operation of an unstaggered MTI is a significant challenge. To stabilize the transmitter sufficiently for pulse-to-pulse or dwell-to-dwell stagger operation is considerably more difficult. Typically, pulse-to-pulse staggering is used with MTI processing, whereas dwell-to-dwell staggering is used with MTD (filter bank) processing.

For many MTI applications pulse-to-pulse or dwell-to-dwell staggering is preferable to scan-to-scan staggering.[†] For example, if a binomial-weighted three-pulse canceler that has 36%-wide rejection notches is employed and if scan-to-scan pulse staggering is used, 36% of the desired targets would be missing on each scan owing to doppler consideration alone. This might be intolerable for some applications.

[†] The choice between pulse-to-pulse staggering and dwell-to-dwell (MTD) operation is a system concept decision—both approaches have their advantages. For example, pulse-to-pulse staggering will not provide canceling of clutter in the ambiguous range intervals. With dwell-to-dwell staggering, an extra transmitter pulse (also known as a fill pulse) will enable canceling of second range interval clutter.

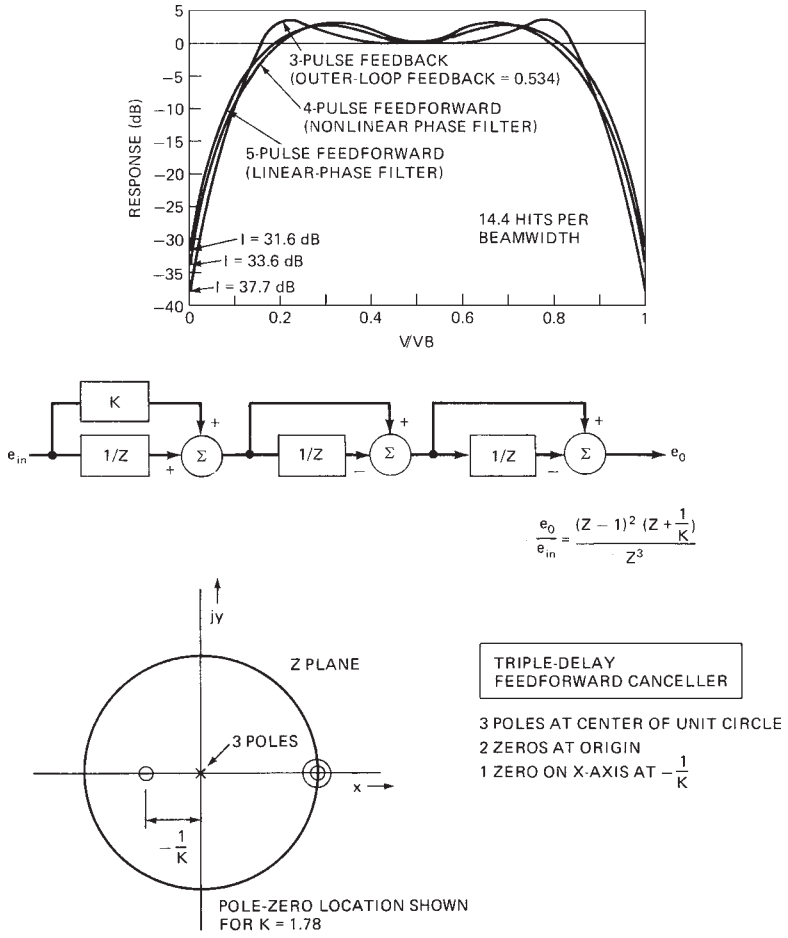


FIGURE 2.38 Shaped-velocity-response feedforward cancelers compared with three-pulse feedback canceler. See text for five-pulse canceler parameters.

With pulse-to-pulse staggering, good response can be obtained on all dopplers of interest on each scan. In addition, better velocity response can be obtained at some dopplers than either pulse interval will give on a scan-to-scan basis. This is so because pulse-to-pulse staggering produces doppler components in the passband of the MTI filter. Pulse-to-pulse staggering may degrade the improvement factor attainable, as shown in Figure 2.23 and Figure 2.24, but this degradation may not be significant, or it can be eliminated by the use of time-varying weights as described below. One further advantage of pulse-to-pulse staggering is that it may permit eliminating the use of feedback in the cancelers (used to narrow the blind-speed notches), which eliminates the transient settling problem of the feedback filters.

The optimum choice of the stagger ratio depends on the velocity range over which there must be no blind speeds and on the permissible depth of the first null

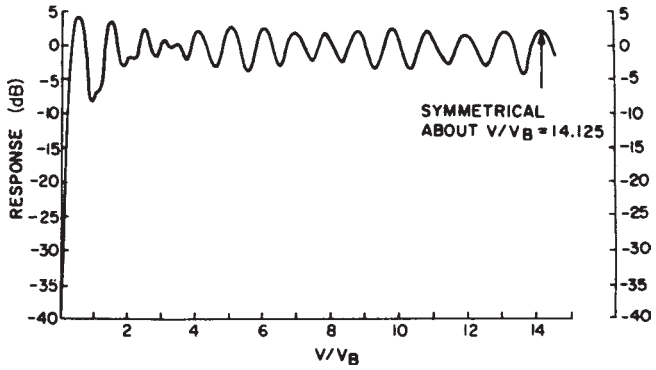


FIGURE 2.39 Velocity response curve: dual canceler, no feedback, 25:30:27:31 pulse-interval ratio

in the velocity response curve. For many applications, a four-period stagger ratio is best, and a good set of stagger ratios can be obtained by adding the first blind speed (in V/V_B) to the numbers −3, 2, −1, 3 (or 3, −2, 1, −3). Thus, in Figure 2.41‡, where the first blind speed occurs at about $V/V_B = 14$, the stagger ratio is 11:16:13:17§ (alternating the long and short periods keeps the transmitter duty cycle as nearly constant as possible, as well as ensuring good response at the first null where $V = V_B$). Figures 2.39 and 2.42 show two other 4-period velocity response curves. If using four interpulse periods makes the first null to be too deep, then five interpulse periods may be used, with the stagger ratio obtained by adding the first blind speed to the number −6, +5, −4, +4, +1. Figure 2.40 shows a velocity response curve for five pulse intervals. The depth of the first null can be predicted from Figure 2.45, which is discussed later.

For a radar system with relatively few hits per beamwidth, it is not advantageous to use more than four or five different intervals because then the response to an individual target will depend on which part of the pulse sequence occurs as the peak of the beam passes the target. Random variation of the pulse intervals is not desirable (unless used as an electronic counter-countermeasure feature) because it permits the nulls to be deeper than the optimum choice of four- or five-pulse intervals.

When the ratio of pulse intervals is expressed as a set of relatively prime integers (i.e., a set of integers with no common divisor other than 1), the first true blind speed occurs at

$$\frac{V}{V_B} = \frac{R_1 + R_2 + R_3 + \cdots + R_N}{N} \tag{2.36}$$

‡ All velocity response curves plotted herein present the average power response of the output pulses of the canceler for the duration of the time on target for a scanning radar. If staggering were used with batch processing, such as in a phased array, these curves would not apply for a single output. For example, if the stagger ratio was 11:16:13:17 and a three-pulse FIR filter is used, it would be necessary to transmit six pulses, with interpulse spacings of 11:16:13:17:11 and sum the power output from the filter after the last four pulses were transmitted to get the equivalent response shown in these curves.

§ Note that the first differences between all combinations of the integers 11, 16, 13, and 17 are 1,2,3,4,5,6. This "perfect difference set" for the stagger sequence is the key to the relative flatness of the response curves.

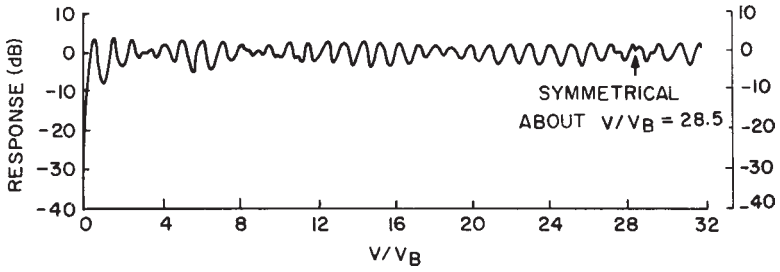


FIGURE 2.40 Velocity response curve: three-pulse binomial canceler, 51:62:53:61:58 pulse-interval ratio

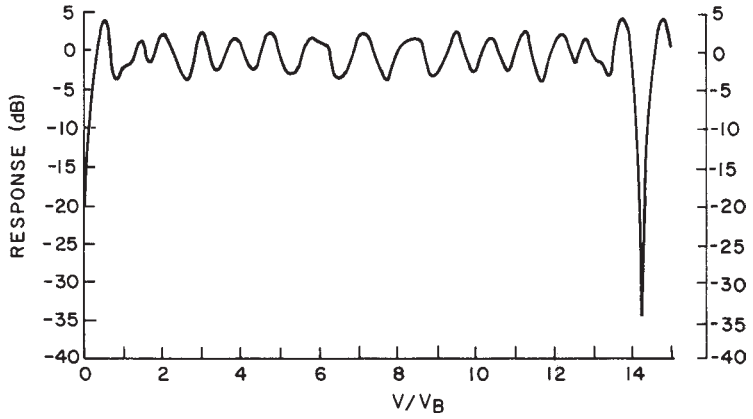


FIGURE 2.41 Velocity response curve: three-pulse binomial canceler, 11:16:13:17 pulse-interval ratio

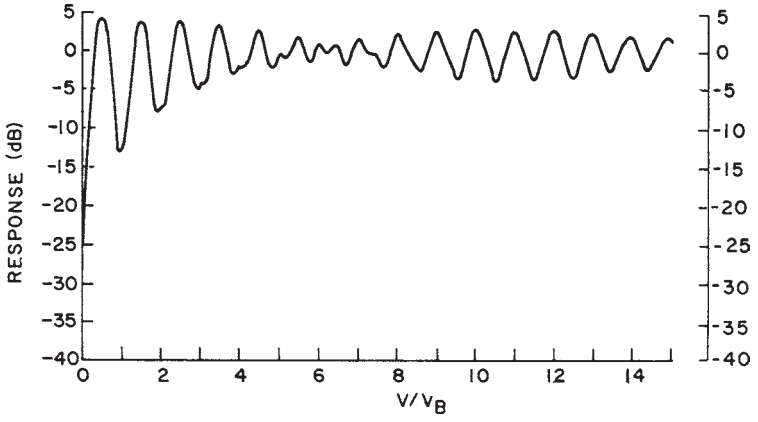


FIGURE 2.42 Velocity response curve: three-pulse binomial canceler, 53:58:55:59 pulse-interval ratio. This response curve continues to $V/V_B = 53$ with no dips below 5 dB. The first blind speed is at $V/V_B = 56.25$.

where $(R_1, R_2, R_3, \ldots, R_N)$ are the set of integers and V_B is the blind speed corresponding to the average interpulse period. The velocity response curve is symmetrical about one-half of the value from Eq. 2.36.

Feedback and Pulse-to-Pulse Staggering. When pulse-to-pulse staggering is employed, the effect of feedback is reduced. Staggering causes a modulation of the signal doppler at or near the maximum response frequency of the canceler. The amount of this modulation is proportional to the absolute target doppler so that, for an aircraft flying at V_B, the canceler response is essentially independent of the feedback employed. Figure 2.43 shows a plot of the effects of feedback on a dual-canceler system with 14.4 hits per beamwidth and a ratio of stagger intervals of 6:7:8. The feedback values employed are several of those used for the unstaggered velocity response plot in Figure 2.35. If scan-to-scan pulse-interval staggering had been used instead of pulse-to-pulse, the no-feedback rms response for three scans at a target velocity of V_B would be −12.5 dB. The composite response for pulse-to-pulse staggering, however, is only −6 dB at V_B, thus illustrating the advantage of pulse-to-pulse staggering.

Improvement Factor Limitations Caused by Staggering. When pulse-to-pulse staggering is used, it limits the attainable improvement factor owing to the unequal time spacing of the received clutter samples. The curves in Figure 2.23 and Figure 2.24, which have been referred to several times, give the approximate limitation on I caused by pulse-to-pulse staggering and either antenna scanning or internal clutter motion. They have been derived as explained below.

A two-delay canceler will perfectly cancel a linear waveform, $V(t) = c + at$, if it is sampled at equal time intervals independent of the constant c or the slope a. (Additional delay cancelers perfectly cancel additional waveform derivatives; e.g., a three-delay canceler will perfectly cancel $V(t) = c + at + bt^2$.) A stagger system with two pulse intervals samples the linear waveform at unequal intervals, and therefore

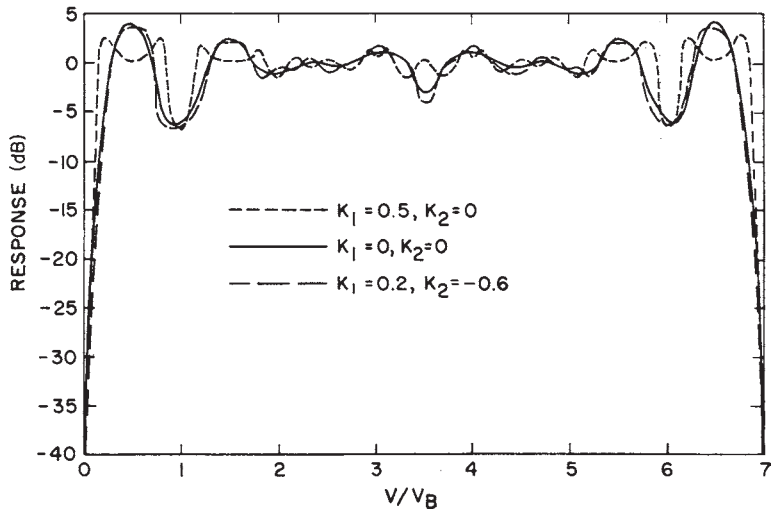


FIGURE 2.43 Effect of feedback on the velocity response curve: dual canceler, 6:7:8 pulse-interval ratio

there will be a voltage residue from the cancelers that is proportional to the slope a and inversely proportional to $\gamma - 1$, where γ is the ratio of the intervals. The apparent doppler frequency of the residue will be at one-half the average repetition rate of the system and thus will be at the frequency of maximum response of a binomial-weight canceler.

The rate of change of phase or amplitude of clutter signals in a scanning radar is inversely proportional to the hits per beamwidth, n. Thus, with the use of a computer simulation to determine the proportionality constant, the limitation on I due to staggering is approximately

$$I \approx 20 \log \left(\frac{2.5n}{\gamma - 1} \right) \text{dB} \tag{2.37}$$

which is plotted in Figure 2.23.

These curves, which apply to all multiple-delay cancelers, give answers that are fairly close to the actual limitation that will be experienced for most practical stagger ratios. An example of the accuracy is as follows: A system with 14.4 hits per beamwidth, a four-pulse binomial weight canceler, and a 6:9:7:8 pulse-interval ratio has an improvement factor limitation of 36.5 dB due to staggering. The curve gives a limitation of 37.2 dB for this case. But if the sequence of pulse intervals were to be changed from 6:9:7:8 to 6:8:9:7, the actual limitation would be 41.1 dB, which is 3.9 dB less than that indicated by the curve. This occurs because the primary modulation with a 6:9:7:8 pulse-interval ratio looks like a target at maximum-response speed, whereas the primary modulation with a 6:8:9:7 pulse-interval ratio looks like a target at one-half the speed of maximum response. Because it is desirable to average the transmitter duty cycle over as short a period as possible, the 6:9:7:8 pulse-interval ratio would probably be chosen for a practical system.

Once Eq. 2.37 for the limitation on I due to scanning and staggering is obtained, it is possible to determine the limitation on I due to internal-clutter motion and staggering. If

$$n = \frac{\sqrt{1n(2)}}{2\pi} \times \frac{\lambda f_r}{\sigma_v} = 0.1325 \frac{\lambda f_r}{\sigma_v} \tag{2.38}$$

(from Eqs. 2.14 and 2.15) is substituted into Eq. 2.37,

$$I = 20 \log \left(\frac{2.5}{\gamma - 1} \times \frac{0.1325 \lambda f_r}{\sigma_v} \right) = 20 \log \left(\frac{0.33 \lambda f_r}{(\gamma - 1)\sigma_v} \right) \tag{2.39}$$

where λ is the wavelength, f_r is the average pulse repetition frequency, and σ_v is the rms velocity spread of scattering elements. This is plotted in Figure 2.24 for rain and for wooded hills with a 40 knot wind. This limitation on the MTI improvement factor is independent of the type of canceler employed.

Time-Varying Weights. The improvement factor limitation caused by pulse-to-pulse staggering can be avoided by the use of time-varying weights in the canceler forward paths instead of binomial weights. The use of time-varying weights has no appreciable effect on the MTI velocity response curve. Whether the added complexity of utilizing time-varying weights is desirable depends on whether the stagger

limitation is predominant. For two-delay cancelers, the stagger limitation is often comparable with the basic canceler capability without staggering. For three-delay cancelers, the stagger limitation usually predominates.

Consider the transmitter pulse train and the canceler configurations shown in Figure 2.44. During the interval T_N when the returns from transmitted pulse P_N are being received, the two-delay canceler weights should be

$$\begin{aligned} A &= 1 \\ C &= \frac{T_{N-2}}{T_{N-1}} \\ B &= -1 - C \end{aligned} \tag{2.40}$$

and the three-delay canceler weights should be

$$\begin{aligned} A &= 1 \\ C &= 1 + \frac{T_{N-3} + T_{N-1}}{T_{N-2}} \\ B &= -C \\ D &= -1 \end{aligned} \tag{2.41}$$

These weights have been derived by assuming that the cancelers should perfectly cancel a linear waveform $V(t) = c + at$, sampled at the stagger rate, independent of the values of the constant c or the slope a. (As mentioned at the beginning of this section, a multiple-delay canceler with binomial weights in an unstaggered system will perfectly cancel $V(t) = c + at$.)

The choice of $A = 1$ in both cases is arbitrary. In the three-delay canceler, setting $D = -1$ eliminates the opportunity for a second-order correction to cancel the quadratic term bt^2, which could be obtained if D were also time-varying. Computer calculations have shown that it is unnecessary to vary D in most practical systems.

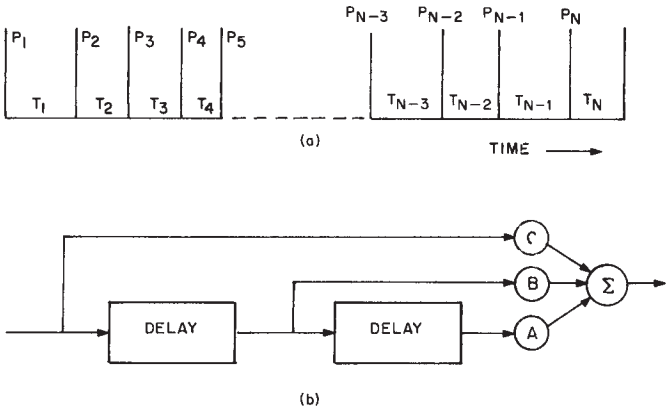


FIGURE 2.44 Use of time-varying weights: (*a*) pulse train, (*b*) two-delay canceler, and (*c*) three-delay canceler

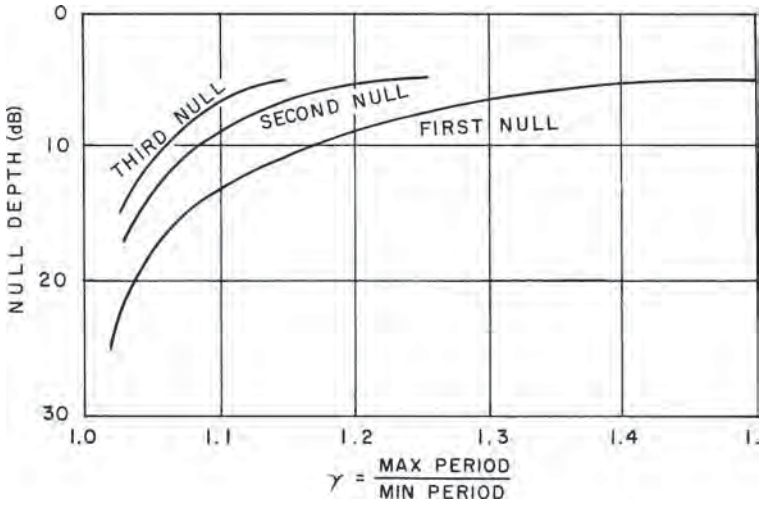


FIGURE 2.45 Approximate depth of nulls in the velocity response curve for pulse-to-pulse staggered MTI

Depth of First Null in Velocity Response. When selecting system parameters, it is useful to know the depth of the first few nulls to be expected in the velocity response curve. As discussed earlier, the null depths are essentially unaffected by feedback. They are also essentially independent of the type of canceler employed, whether single, dual, or triple, or of the number of hits per beamwidth. Figure 2.45 shows approximately what null depths can be expected versus the ratio of maximum to minimum interpulse period.

2.9 MTI FILTER DESIGN FOR WEATHER RADARS

MTI filters are used at the lower elevation angles in weather radars to prevent weather estimates from being contaminated with ground clutter returns. It is, however, also very important to preserve an accurate measurement of weather intensity and precipitation rate. To meet this dual objective, MTI filters with narrow fixed clutter rejection notches and flat passbands are needed. Use of a very narrow clutter notch even permits measuring weather precipitation rates with a mean radial velocity of zero, albeit with some bias.* Such measurement is possible because weather usually has a wide spectral spread—typically 1 to 4 m/s—whereas fixed clutter has a much narrower spectral spread—typically less than 0.5 m/s.

* *Bias* as used herein refers to the error in measuring radar reflectivity due to the clutter notch and lack of flatness of the MTI filters. When weather has a wide spectral spread and the clutter notch of the filters is narrow, there is minimal measurement error induced by the MTI filters. Conversely, when the weather spectral width is narrow and the radial velocity of the weather is near zero, significant error in the weather reflectivity measurement will exist. There are other causes of error between radar estimates of precipitation rates and rain gauge measurements that are not addressed herein, such as the spatial and temporal distribution of rain.

Examples of weather radar applications for which MTI filters are used:

1. *Weather Doppler Radars (NEXRAD/WSR-88)* Radars with rotating antennas that measure precipitation rate, doppler velocity, and turbulence. Measures total rainfall and provides tornado warnings.
2. *Terminal Doppler Weather Radars (TDWR)* Radars with rotating antennas designed to detect severe wind shear in aircraft approach and departure paths close to airports.
3. *Airport Surveillance Radars* Radars with rotating antennas designed for air traffic control functions in the terminal area but with a secondary function of detecting and monitoring severe weather and wind shear in aircraft approach and departure paths.
4. *Phased Array Radars* Radars with fixed electronically scanned antennas designed for many functions such as missile detection and air traffic control, and used concurrently for measuring precipitation rates.

As an example, the design of elliptic MTI filters as used in the TDWR will be described. TDWR is a C-band radar used at airports for detection of downbursts, microbursts, and prediction of wind direction. Elliptic filters are infinite impulse response (IIR) filters that have the sharpest possible transition from rejection notch to passband for a specified level of the clutter rejection notch (width and depth), ripple in the passband, and number of delay sections (see Oppenheim and Schafer[28]). The elliptic filters can be followed with pulse-pair processing[13] for estimation of weather mean velocity and spectral width (turbulence). There are two drawbacks of elliptic filters: First, the long transient settling time. For a scanning weather radar, it takes about four beamwidths of scanning after the transmitter starts pulsing before clutter attenuation reaches 50 to 60 dB. Second, if the input clutter signal reaches the limit level in the IF receiver, there will be a significant transient increase of clutter residue. One of the elliptic filters employed in the original TDWR radar is used as an example.

TDWR operates at C band (5.60-5.65 GHz). The antenna rotates at 4.33 rpm and has a 0.55° one-way beamwidth. The PRF is 1066 Hz. The elliptic filter designed for these parameters has an improvement factor of 57.2 dB. HBW (hits per one-way 3-dB beamwidth) are 22.6. The specifications for the elliptic filter for the above parameters are normalized stopband edge $\sigma_f T = 0.03492$; passband edge $\sigma_f T = 0.07350$; stopband attenuation 58 dB below peak filter response; and passband ripple = 2.0 dB. To meet these requirements, the filter requires 4-delay sections, which can be implemented as two cascaded 2-delay sections, as shown in Figure 2.46.

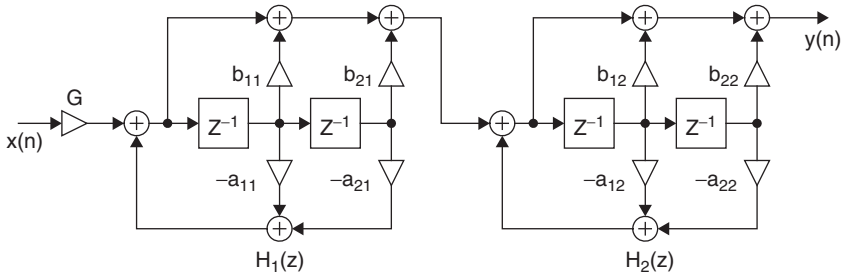


FIGURE 2.46 Four-delay elliptic filter used in TDWR

The filter coefficients are

$$a_{11} = -0.901933 \qquad a_{12} = -1.701983$$
$$a_{21} = 0.420985 \qquad a_{22} = 0.914913$$
$$b_{11} = 1.000000 \qquad b_{21} = 1.000000$$
$$b_{21} = -1.992132 \qquad b_{22} = -1.958290$$

The calculated improvement factor for this filter against land clutter with 22.6 HBW is 58 dB, and the bias for weather returns with spectral spreads of 1 and 4 m/sec is –10 dB and –2 dB, respectively, when the radial velocity of the weather returns is $v = 0$ m/s.

Figure 2.47 shows the elliptic filter CW response and its response for weather with 1 m/s and 4 m/s rms spectral spread. The unambiguous doppler interval corresponding to $f_d T = 1$ is 28.4 m/s for the parameters used to calculate this response.

Figure 2.48 shows the time-domain responses for this filter as the antenna scans past a point of clutter, such as a water tower. This figure shows the input to the elliptic filter and the residue output. A gaussian antenna pattern is assumed in this figure. The calculated improvement factor for the sequence shown (total clutter power into the filter divided by total residue power out of the filter, normalized by the noise gain of the filter) is 58 dB.

A $\sin(x)/x$ antenna pattern is assumed for the following three figures, but the lessons to be gained from these figures is essentially independent of the assumed beam shape. Figure 2.49 shows the filter response if the transmitter starts radiating just as a null of the antenna pattern passes the point of clutter. The individual samples of residue are 60 or more dB below the peak clutter return. The improvement factor for this sequence is 57 dB.

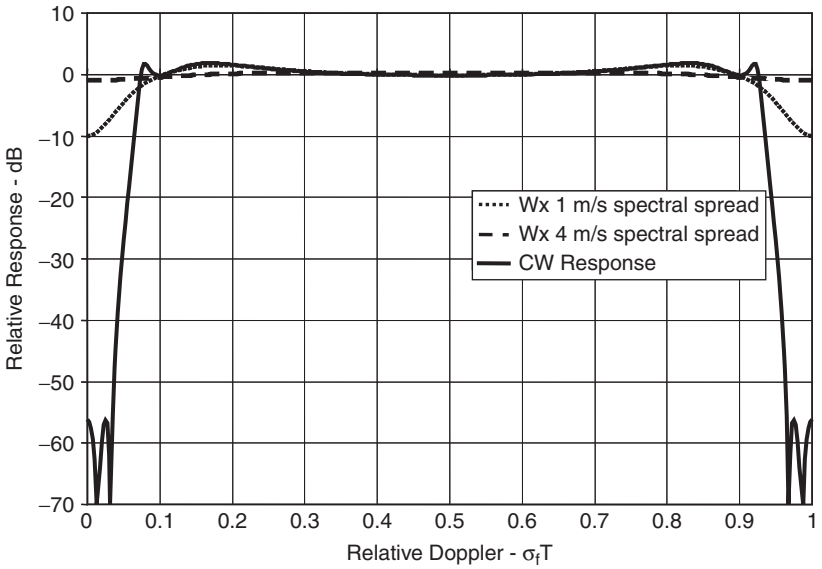


FIGURE 2.47 Elliptic filter CW response and response to weather with $\sigma = 1$ and 4 m/s rms spectral spread

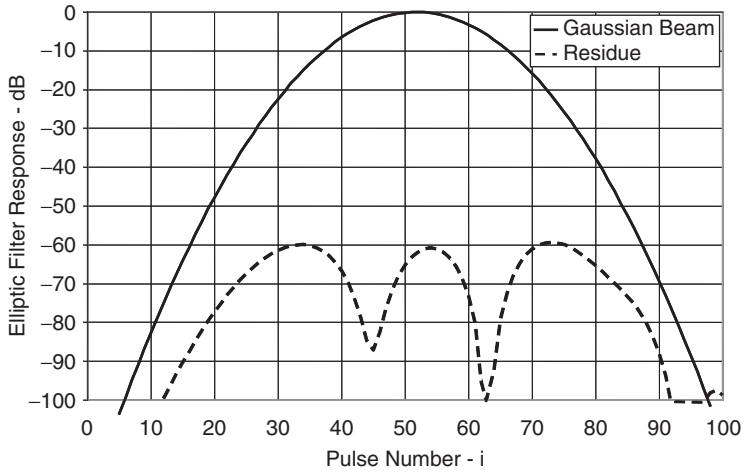


FIGURE 2.48 Time-domain clutter input and output residue as antenna scans past a point target

Figure 2.50 shows the residue if the transmitter starts radiating as the peak of the beam passes the point clutter. Forty-nine pulses after the transmitter starts radiating, the residue has decayed only 27 dB. It would take at least another 50 pulses for the residue to decay to −60 dB. For this reason, when the transmitter starts pulsing, a settling time of at least 90 pulses must be allowed before useful data is collected.

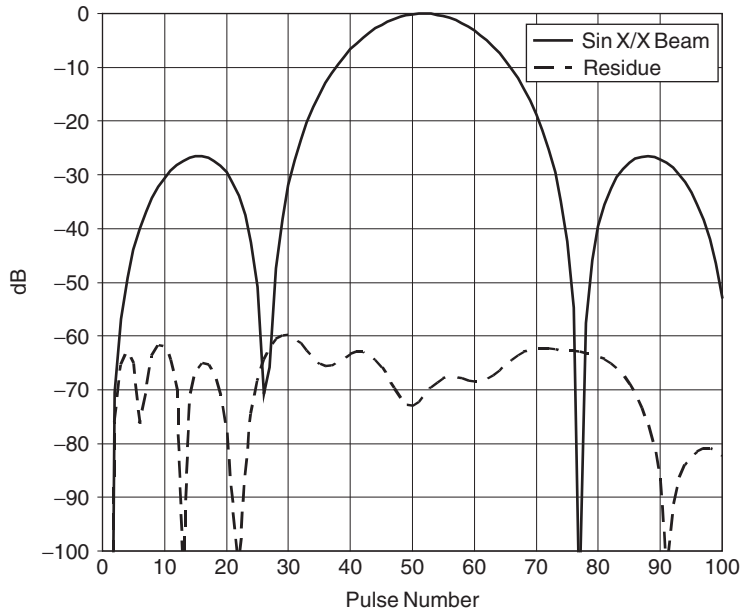


FIGURE 2.49 Clutter input and residue from elliptic filter. Radar starts radiating at pulse number 1.

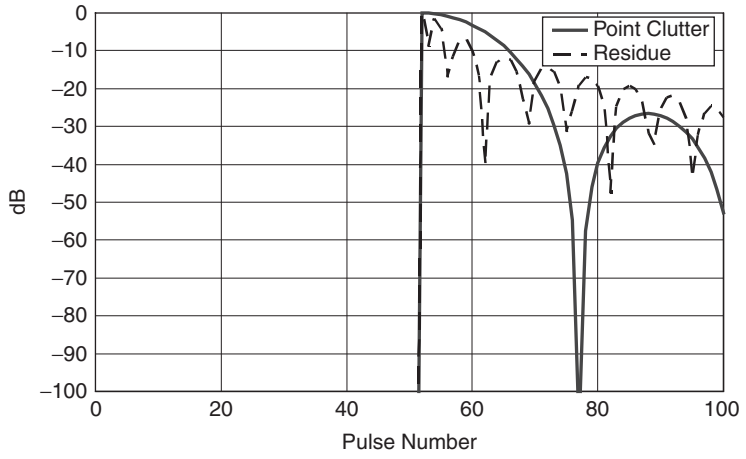


FIGURE 2.50 Clutter input and residue from elliptic filter. Radar starts radiating at pulse number 51.

Figure 2.51 shows the effect of the returned signal if the point clutter exceeds the IF limit level by 6 dB. When the signal reaches the limit level, there is a step increase of residue of about 30 dB. TDWR uses clutter maps to normalize the residue from the strong points of clutter that exceed the limit level.

The weather mode of Airport Surveillance Radars is demonstrated by five-pulse finite impulse response (FIR) filters used in the ASR-11, an S-band radar used for air traffic control at airports. The design of the filters is primarily for Moving Target Detector (MTD) detection of aircraft, but special attention is given to providing flat passband response for accurate weather reflectivity estimation. The filter bank (for HBW = 17) is pictured in Figure 2.52 and the coefficients are shown in Table 2.3.

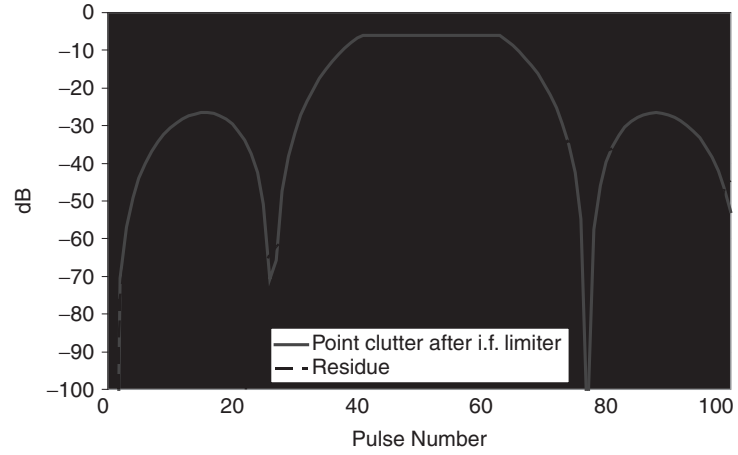


FIGURE 2.51 Effect of limiting on elliptic filter response

TABLE 2.3 ASR: Coefficients of ASR-11 5-Pulse Low-PRF Fir Filters

FILTER	Coefficient 1	Coefficient 2	Coefficient 3	Coefficient 4	Coefficient 5
20 dB	0.79812	–0.50687	–0.29297	–0.08340	0.11528
30 dB	0.67844	–0.62907	–0.28700	0.00815	0.24810
40 dB	0.50178	–0.80291	0.06899	0.30685	–0.06807
50 dB	0.39235	–0.78485	0.21613	0.37851	–0.20021
60 dB	0.28502	–0.75401	0.58529	–0.03661	–0.07956
70 dB	0.17766	–0.58440	0.70278	–0.35920	0.06322

Selection of filters is based on clutter amplitude information stored in a clutter map. The filters are selected on a range-cell by CPI basis.

These FIR clutter filters have the narrowest rejection notches that can be obtained with five pulses and the indicated level of fixed clutter rejection. However, the notches are significantly wider than those of the elliptic filters; thus, they will have greater bias for measurement of weather intensity when the weather radial velocity is zero.

For phased array radars, FIR filters similar to those described for the ASR-11 are applicable. The filters can be designed, if the time budget of the phased array radar allows, to utilize more than the five pulses per coherent processing interval (CPI) used by the ASR-11 radar. Using more pulses makes possible narrower rejection notches and thus less bias for estimates of precipitation with zero radial velocity.

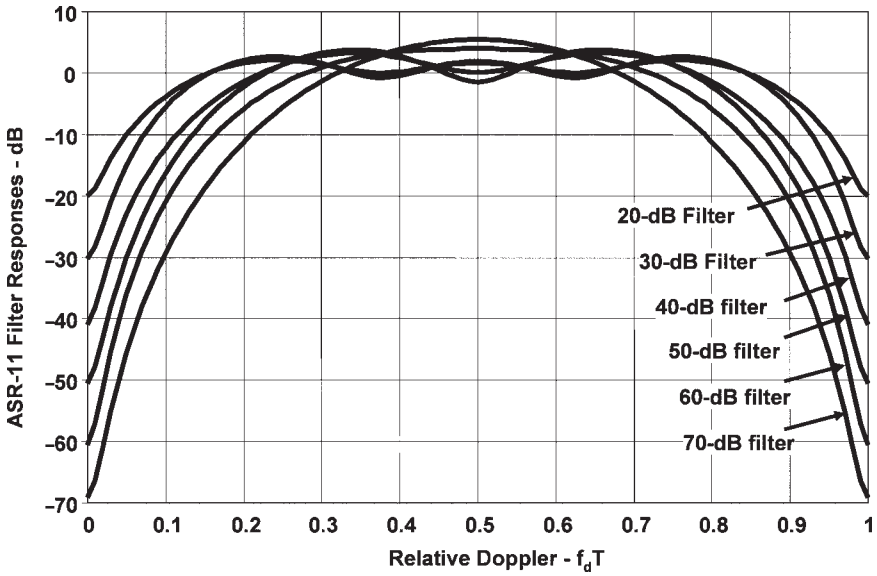


FIGURE 2.52 Response of ASR-11 FIR filters low-PRF ($f_r = 855$ pps) filters operating against fixed clutter with HBW = 17. The unambiguous doppler interval ($fT = 1$) is 45.8 m/s for the parameters used to calculate this response.

2.10 CLUTTER FILTER BANK DESIGN

As discussed in Section 2.2, the MTD uses a waveform consisting of coherent processing intervals (CPIs) of N pulses, all at the same PRF and RF frequency. The PRF and possibly the RF are changed from one CPI to the next. With this constraint, only finite-impulse-response (FIR) filter designs are realistic candidates for the filter bank design. (Feedback filters require a number of pulses to settle after either the PRF or the RF is changed and thus would not be practical.)

The number of pulses available during the time when a surveillance radar beam illuminates a potential target position is determined by system parameters and requirements such as beamwidth, PRF, volume to be scanned, and the required data update rate. Given the constraint on the number of pulses on target, one must decide how many CPIs should occur during the time on target and how many pulses per CPI. The compromise is usually difficult. One wishes to use more pulses per CPI to enable the use of better filters, but one also wishes to have as many CPIs as possible. Multiple CPIs (at different PRFs and perhaps at different RF frequencies) improve detection and can provide information for true radial velocity determination.[30]

The design of the individual filters in the doppler filter bank is a compromise between the frequency sidelobe requirement and the degradation in the coherent integration gain of the filter. The number of doppler filters required for a given length of the CPI must be balanced between hardware complexity and the straddling loss at the crossover between filters. Finally, the requirement of providing a high degree of clutter suppression at zero doppler (land clutter) sometimes introduces special design constraints.

When the number of pulses in a CPI is large (≥ 16), the systematic design procedure and efficient implementation of the fast Fourier transform (FFT) algorithm is particularly attractive. Through the use of appropriate weighting functions of the time-domain returns in a single CPI, the resulting frequency sidelobes can be readily controlled. Further, the number of filters (equal to the order of the transform) needed to cover the total doppler space (equal to the radar PRF) can be chosen independently of the CPI, as discussed below.

As the CPI becomes smaller (≤ 10), it becomes important to consider special designs of the individual filters to match the specific clutter suppression requirements at different doppler frequencies in order to achieve better overall performance. While some systematic procedures are available for designing FIR filters subject to specific passband and stopband constraints, the straightforward approach for small CPIs is to use an empirical approach in which the zeros of each filter are adjusted until the desired response is obtained. An example of such filter designs is presented next.

Empirical Filter Design. An example of an empirical filter design for a six-pulse CPI follows. (The six pulses per CPI may be driven by system considerations, such as time-on-target.) Because the filter will use six pulses, only five zeros are available for the filter design; the number of zeros available is the number of pulses minus one. The filter design process consists of placing the zeros to obtain a filter bank response that conforms to the specified constraints. The example that follows was produced with an interactive computer program with which the zeros could be moved until the desired response was obtained. The assumed filter requirements are as follows:

- Provide a response of −66 dB in the clutter rejection notch (relative to the peak target response) of the moving-target filters.

- Provide a response of −46 dB for chaff rejection at velocities between ±20% of the ambiguous doppler frequency range.
- In this design, only five filters will be implemented.
- Three of the five filters will reject fixed clutter and respond to moving targets. Two filters will respond to targets at zero doppler and its ambiguities. (With good fixed clutter rejection filters, it takes two or more coherent filters to cover the gap in response at zero velocity.)

With the above considerations, a filter bank can be constructed.

Figure 2.53*a* shows the filter designed to respond to targets in the middle of the doppler passband. The sidelobes near zero velocity are 66 dB down from the peak, thus providing good clutter rejection for clutter within 5% of zero doppler. The −46 dB sidelobe provides chaff rejection to ±16%. Because of the constraint of having only five zeros available, this filter could not provide −46 dB rejection to ±20%.

Figure 2.53*b* shows the filter that responds to targets as near as possible to zero doppler, while having a zero-doppler response of −66 dB. Two zeros are placed near 0, providing −66 dB response to clutter at 0. The filter sidelobes between 0.8 and 1.0 doppler provide the specified chaff rejection of 46 dB. A mirror image of this filter is used for the third moving doppler filter. (The mirror-image filter has coefficients that are complex conjugates of the original filter coefficients.)

Figure 2.53*c* shows the first filter designed for response at zero doppler. Considerations here are that the doppler straddling loss of the filter bank be minimized

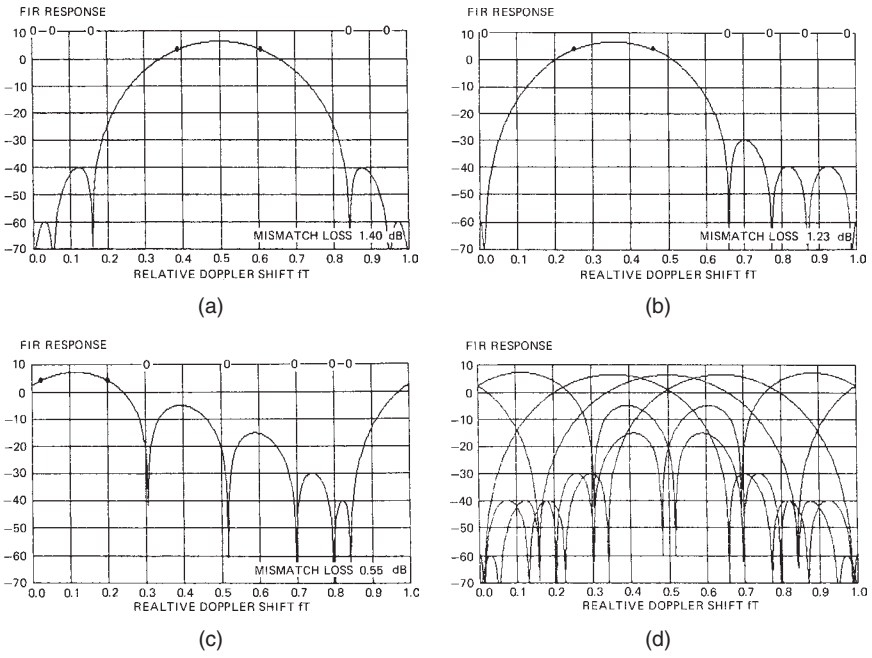


FIGURE 2.53 Six-pulse filters for targets at (*a*) $fT = 0.5$, (*b*) $Ft = 0.3$, $fT = 0.8$, and (*c*) combined response of complete bank of five six-pulse filters

(this dictates the location of the peak), that the response to chaff at 0.8 doppler be down 46 dB, and that the mismatch loss be minimized. Minimizing the mismatch loss is accomplished by permitting the filter sidelobes between 0.3 and 0.8 to rise as high as needed (lower sidelobes in this range increase the mismatch loss). The second zero-doppler filter is the mirror image of this one.

Figure 2.53*d* shows the composite response of the filter bank. Note that the filter peaks are fairly evenly distributed. The dip between the first zero-doppler filter and the first moving doppler filter is larger than the others, primarily because, under the constraints, it is impossible to move the first doppler filter nearer to zero velocity.

Chebyshev Filter Bank. For a larger number of pulses in the CPI, a more systematic approach to filter design is desirable. If a doppler filter design criterion is chosen that requires the filter sidelobes outside the main response to be below a specified level (i.e., providing a constant level of clutter suppression), while simultaneously minimizing the width of the filter response, a filter design based on the Dolph-Chebyshev distribution provides the optimum solution. Properties and design procedures based on the Dolph-Chebyshev distribution can be found in the antenna literature. An example of a filter design for a CPI of 9 pulses and a sidelobe requirement of 68 dB is shown in Figure 2.54. The peak filter response can be located arbitrarily in frequency by adding a linear-phase term to the filter coefficients.

The total number of filters implemented to cover all doppler frequencies is a design option trading straddling loss at the filter crossover frequencies against implementation complexity. An example of a complete doppler filter bank implemented with nine uniformly spaced filters is shown in Figure 2.55. The performance of this doppler filter bank against the clutter model considered in Figure 2.25 is shown in Figure 2.56. This graph shows the signal-to-clutter ratio improvement against clutter at zero doppler as a function of target doppler frequency. Only the response of the filter providing the greatest improvement is plotted at each target doppler.

For comparison the optimum curve from Figure 2.25 is shown by a broken line and thus provides a direct assessment of how well the Chebyshev filter design performs against a given clutter model. Also shown is the average SCR improvement for both the optimum and the Chebyshev filter bank.

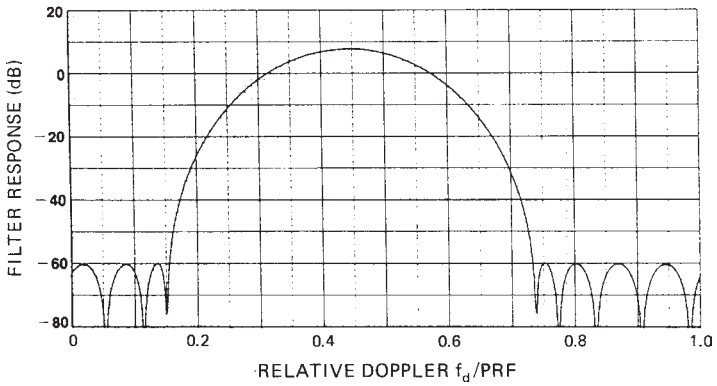


FIGURE 2.54 Chebyshev FIR filter design with 68 dB doppler sidelobes

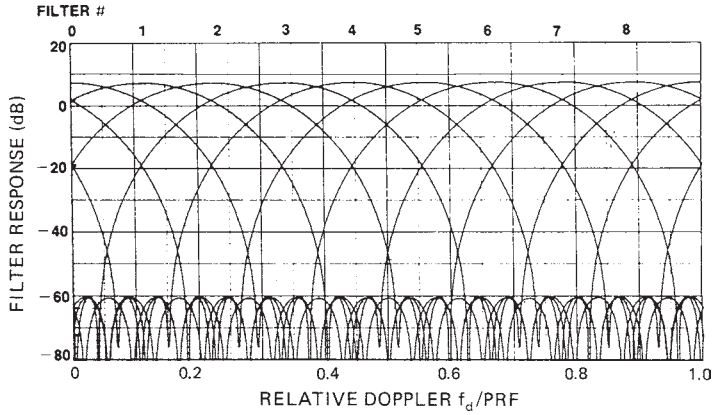


FIGURE 2.55 Doppler filter bank of 68 dB Chebyshev filters, CPI = 9 pulses

Finally, Figure 2.57 shows the average SCR improvement of the 68 dB Chebyshev doppler filter bank as well as the optimum curve (from Figure 2.26) as a function of the relative spectrum spread of the clutter. Owing to the finite number of filters implemented in the filter bank, the average SCR improvement will change by a small amount if a doppler shift is introduced into the clutter returns. This effect is illustrated by the cross-hatched region, which shows upper and lower limits on the average SCR improvement for all possible clutter doppler shifts. For a smaller number of filters in the doppler filter bank, this variation would be greater.

Fast Fourier Transform Filter Bank. For a large number of parallel doppler filters, hardware implementation can be simplified significantly through the use of the FFT algorithm. The use of this algorithm constrains all filters in the filter bank to

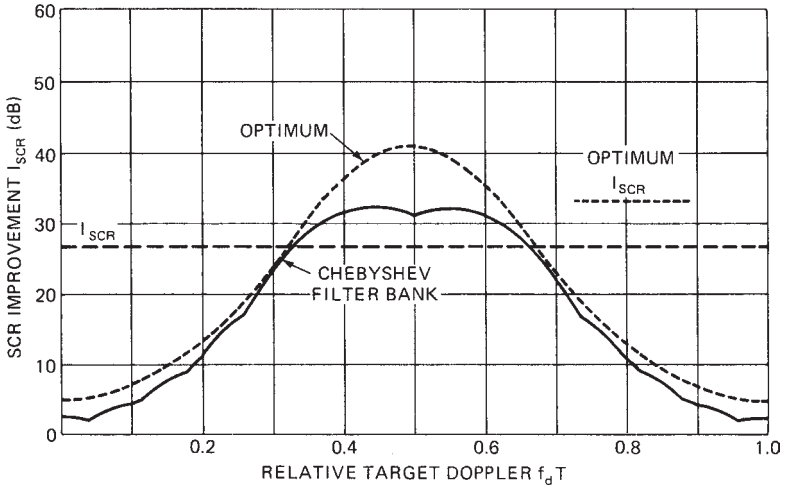


FIGURE 2.56 SCR improvement of 68 dB Chebyshev doppler filter bank compared with the optimum

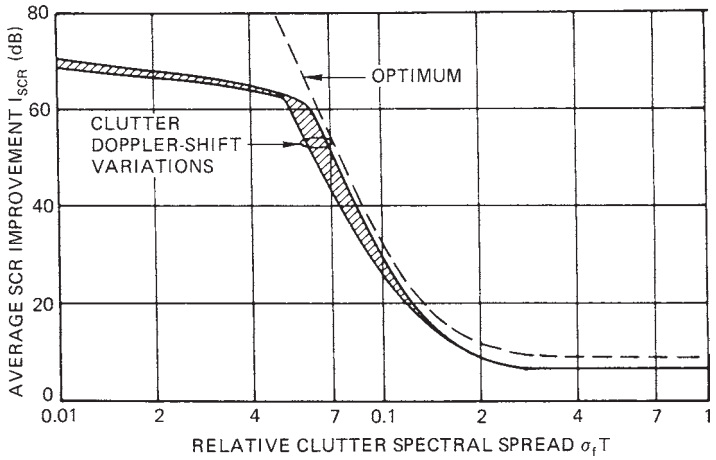


FIGURE 2.57 Average SCR improvement for the 68 dB Chebyshev filter bank shown in Figure 2.55. CPI = 9 pulses. Optimum is from Figure 2.26.

have identical responses, and the filters will be uniformly spaced along the doppler axis. The number of filters implemented for a given size of the CPI can, however, be varied. For example, a greater number of filters can be realized by extending the received data with extra zero values (also known as zero padding) after the received returns have been appropriately weighted in accordance with the desired filter response (e.g., Chebyshev).

Filter Bank Designs Using Constrained Optimization Techniques. For a greater numbers of pulses in the CPI, and when the economy of the FFT implementation of a doppler filter bank can be replaced by a FIR implementation, more desirable FIR filter responses can be realized through the use of appropriate numerical digital filter design techniques. The goal is similar to that pursued with the empirical filter designs discussed earlier but filters with a large number of taps can be designed to exacting specifications.

As an example, consider the design of a doppler filter bank for an S band (3.0 GHz) radar using a CPI of $N = 25$ pulses using a PRF of 6 kHz. Assume that the radar requirements call for a suppression of stationary land clutter by 80 dB and a suppression of moving clutter (rain) by 50 dB. For the filter design, a clutter attenuation 10 dB below these requirements will be needed to keep the sensitivity loss due to the clutter residue below 1 dB and also because each doppler filter will have a coherent gain of around $10 \cdot \log_{10}(25) = 14$ dB, this must be added to the filter design specification as well. The total S-band doppler space for the above radar parameters is 300 m/s, and assuming that the land clutter suppression region has to be ± 4 m/s and that the moving clutter suppression region has to be ± 30 m/s, the constraint for all doppler filter designs normalized to their peak is as shown in Figure 2.58.

Using a signal processing toolbox developed by Dr. Dan P. Scholnik of the Naval Research Laboratory, a doppler filter bank meeting the above constraints was designed. The first filter, which has its peak located as close as possible to the left edge of the constraint box is shown in Figure 2.59, with the abscissa normalized to the total available doppler space.

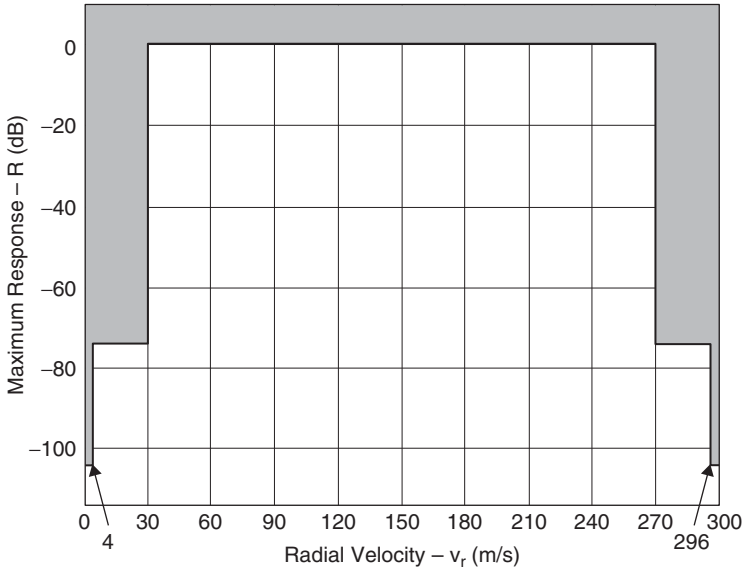


FIGURE 2.58 Doppler filter design constraints

The mismatch loss of this filter is $L_m = 1.29$ dB, which is well below that of a 105 dB Dolph-Chebyshev filter bank ($L_m = 3.0$ dB). For the remaining filters, a relative spacing of $D = 1/25 = 0.04$ was used, but this could be reduced in order to minimize doppler straddling losses. The third filter in the filter bank is shown in Figure 2.60.

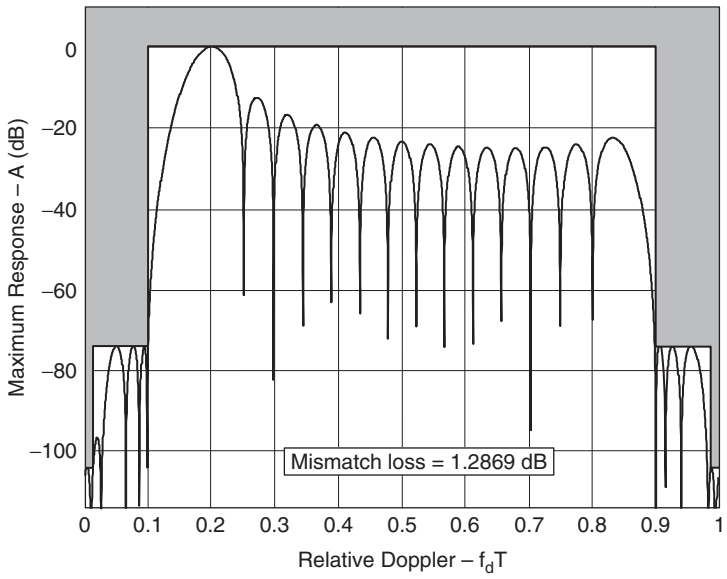


FIGURE 2.59 Leftmost FIR filter in doppler filter bank design

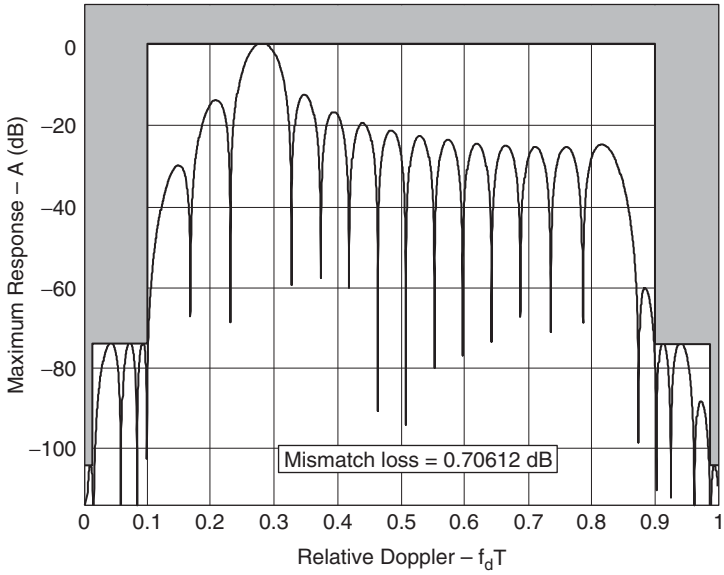


FIGURE 2.60 Third FIR filter in doppler filter bank design

The mismatch loss has now been reduced to 0.71 dB. Finally, the complete doppler filter bank is shown in Figure 2.61. This filter bank could be augmented with additional filters around zero doppler, but these would not meet the design constraints discussed above. The main benefit of a customized doppler filter bank design, as

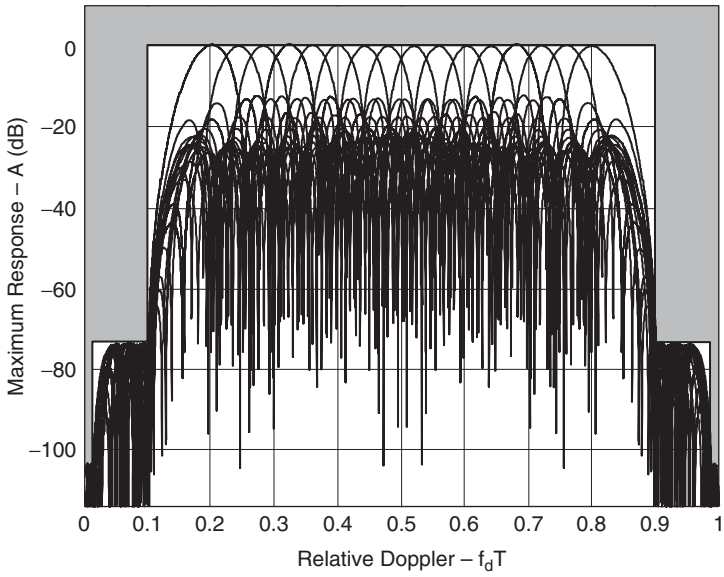


FIGURE 2.61 Complete doppler filter bank design

described here, is its reduced mismatch loss. For the 16 filters in the above design, the average mismatch loss is $\bar{L}_m = 0.66$ dB, a savings of 2.3 dB as compared to the alternative of a 105 dB weighted Dolph-Chebyshev filter bank.

2.11 PERFORMANCE DEGRADATION CAUSED BY RECEIVER LIMITING

Elsewhere in this chapter (Sections 2.2 and 2.12, particularly) IF bandpass limiters have been discussed as (1) a means of preventing received clutter signals from exceeding the range of the A/D converters, (2) normalizing MTI clutter residue caused by system instabilities, and (3) normalizing residue due to the spectral spread of "fixed clutter" caused by either scanning or wind-blown motion. There are occasional clutter residue spikes when clutter exceeds the limit level, and in the past, the energy from these spikes of residue has been suppressed by further reduction of the limit level. When limiters have been used to normalize the energy of clutter residue spikes, the average improvement factor of the MTI systems drastically deteriorates. The equations for I (improvement factor) of a scanning radar in Section 2.6 are based on linear theory. Field measurements, however, have shown that many scanning multiple-delay MTI radar systems fall considerably short of the predicted performance. This occurs because the IF bandpass limiters have been used to suppress the energy of the residue spikes that are caused by the limiting action. Later in this section, it is shown that the use of a binary detection scheme, instead of a drastic reduction of the limit level, can be used to maintain a clutter rejection performance close to linear theory prediction in the resolution cells where clutter limiting occur.

An example of how limiting the dynamic range adjusts the residue is shown in the MTI PPI photographs shown in Figure 2.62. The range rings are at 5-mi intervals.

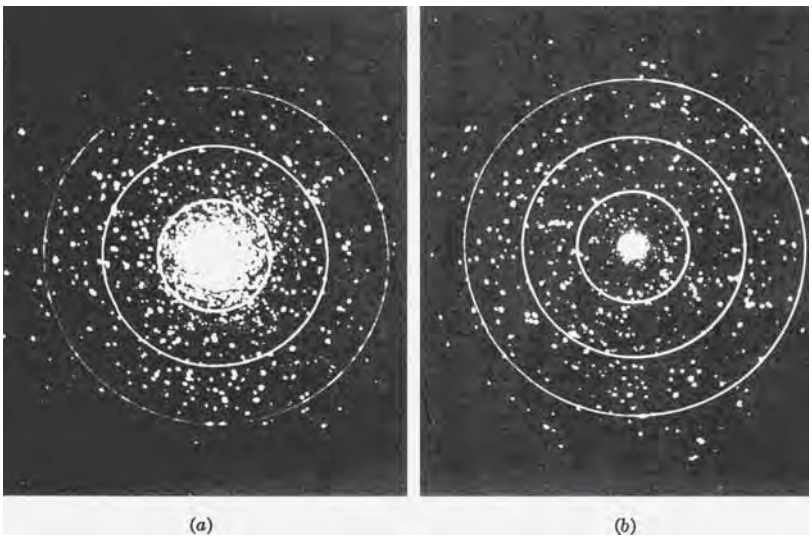


FIGURE 2.62 Effect of limiters: (*a*) 18 dB improvement factor, 20 dB input dynamic range, and (*b*) 18 dB improvement factor, 14 dB input dynamic range

A number of birds are shown on the display. The residue from clutter in the left photograph is solid out to 3 nmi and then decreases until it is almost entirely gone at 10 nmi. The MTI improvement factor in both pictures is 18 dB, but the input dynamic range (peak signal-to-rms noise) to the canceler was changed from 20 to 14 dB between the two pictures. An aircraft flying over the clutter in the first 5 mi in the left-hand picture could not be detected, no matter how large its radar cross section. In the right-hand picture, the aircraft could be detected if the target-to-clutter cross-section ratio were sufficient. Although this example is from many years ago,[31] the principle is still the same, even though current MTI improvement factors are better by tens of dBs. Restriction of the IF dynamic range is still a very efficient way of normalizing clutter residue due to system instabilities or clutter spectral spread to system noise. This is true whether or not the radar uses pulse compression.

Prior to the development of modern clutter maps for controlling false alarms caused by clutter residue, or the more recent suggestion that binary integration can mitigate impulse-like residue,[32] the use of IF limiting was essential for false-alarm control in an MTI radar. Such limiting, however, seriously affects the mean improvement factor obtainable with a scanning-limited, multiple-delay canceler because of the increased spectral spread of the clutter that exceeds the limit level. Part of the additional clutter spectral components comes from the sharp discontinuity in the envelope of returns as the clutter reaches the limit level.[33] A time-domain example of this phenomenon is shown in Figure 2.63 for a radar with $N = 16.4$ hits per beamwidth. On the left is a point target that does not exceed the limit level; on the right is a point target that exceeds the limit level by 20 dB. Note that, for this example, I degrades by 12.8 dB for the dual canceler and by 26.5 dB for the triple canceler. The exact result of this calculation depends on the assumed shape of the antenna pattern; for this example, a $\frac{\sin(u)}{u}$ pattern terminated at the first nulls was assumed. There is a comparable improvement factor degradation due to spectral spreading of limited distributed clutter.[34,35] Figures 2.64, 2.65, and 2.66 show the expected mean improvement factor for two- three-, and four-pulse cancelers as a function of σ/L, the ratio of the rms clutter amplitude to the limit level. Hits per one-way half-power beamwidth are indicated by N.

An example of clutter residue from simulated hard-limited distributed clutter is taken from Hall and Shrader.[32] Figure 2.67 shows a polar plot of part of a linear clutter sequence for a scanning radar with $N = 20$ hits per beamwidth. This linear clutter sequence is 65 consecutive complex voltage returns from one range cell of distributed clutter. Figure 2.68 shows the phase and amplitude of this sequence.

If this clutter sequence were 40 dB stronger and passed through a 10 V IF limiter, only the phase information would remain. Each pulse would have a 10 V amplitude. When the resulting limited clutter sequence is passed through a three-pulse canceler (coefficients 1, –2, 1), the output residue appears as in Figure 2.69*a*. The corresponding pulse-to-pulse improvement factor is shown Figure 2.69*b*.

The expected three-pulse canceler improvement factor (from equation 2.27) for a linear system with $N = 20$ is $I_3 = n^4/3.84 = 46.2$ dB. In Figure 2.69*b,* it is seen that this level of I_3 is achieved for most of the pulses, with only two pulses having very low values of I_3. The statistics for the distribution of I_3 for the three-pulse canceler for hard-limited distributed clutter are shown in Figure 2.70.[32]

Note that for $N = 20$, less that 5% of the hard-limited samples have an improvement factor less than 24 dB, whereas almost 60% of the samples exceed the I_3 expected for a linear system.

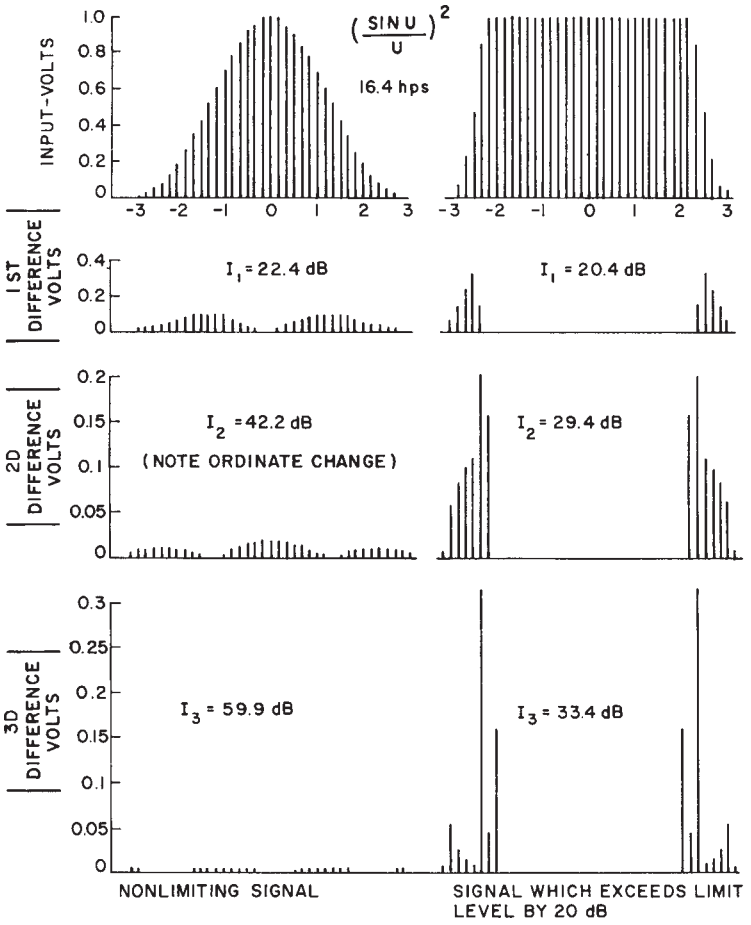


FIGURE 2.63 Improvement factor restriction caused by a limiter

The time-domain illustration shown previously in Figure 2.69 leads to the conclusion of Hall and Shrader[32] that using an *M* out of *N* binary detector at the output of an MTI filter will preclude false alarms from the clutter residues caused by limiting.

Figure 2.71 shows, in addition to clutter residue, the returns from a target that was superimposed on the distributed clutter prior to the clutter-plus-target sequence passing through the IF limiting process. One can see that many of the individual pulse returns from the target exceed the detection threshold, whereas only four of the clutter residue pulses exceed the threshold.

To summarize: (1) The MTI improvement factor in a majority of limiting clutter cells exceeds the average improvement factor obtained with linear processing; (2) cells with poor MTI improvement factor can be rejected with binary detection processing; and, therefore, (3) excellent MTI performance can be obtained even in regions of clutter that exceed the IF dynamic range.

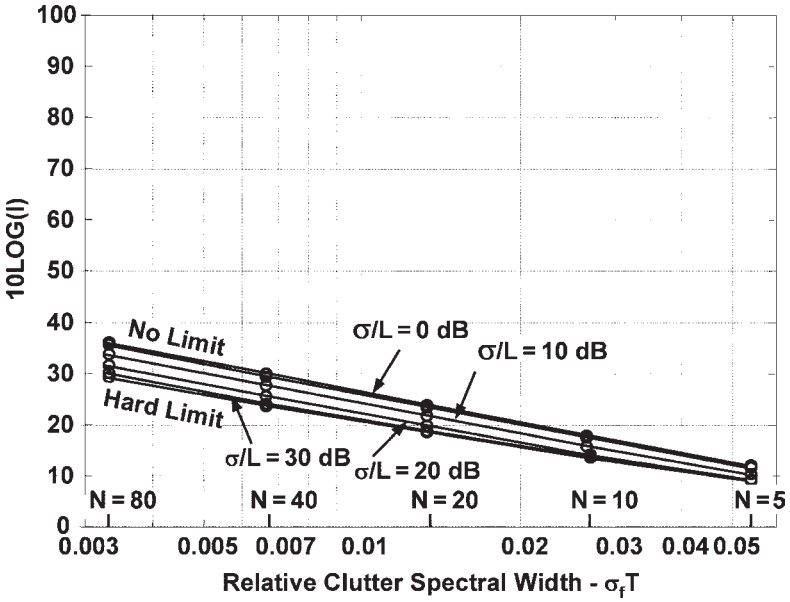


FIGURE 2.64 Mean improvement factor restriction versus amount of limiting and clutter spectral spread for a two-pulse canceler (*after T. M. Hall and W. W. Shrader[32] © IEEE 2007 and H. R. Ward and W. W. Shrader[34] © IEEE 1968*)

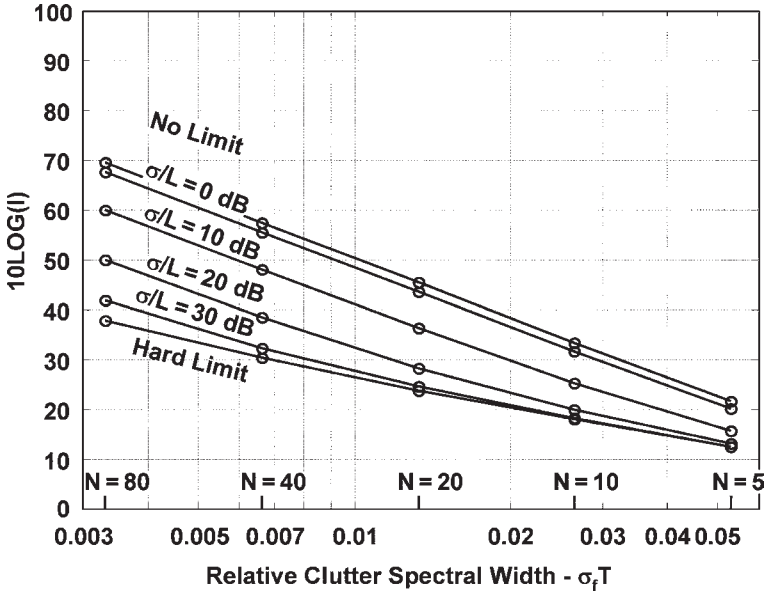


FIGURE 2.65 Mean improvement factor restriction versus amount of limiting and clutter spectral spread for a three-pulse canceler *(after T. M. Hall and W. W. Shrader[32] © IEEE 2007 and H. R. Ward and W. W. Shrader[34] © IEEE 1968*)

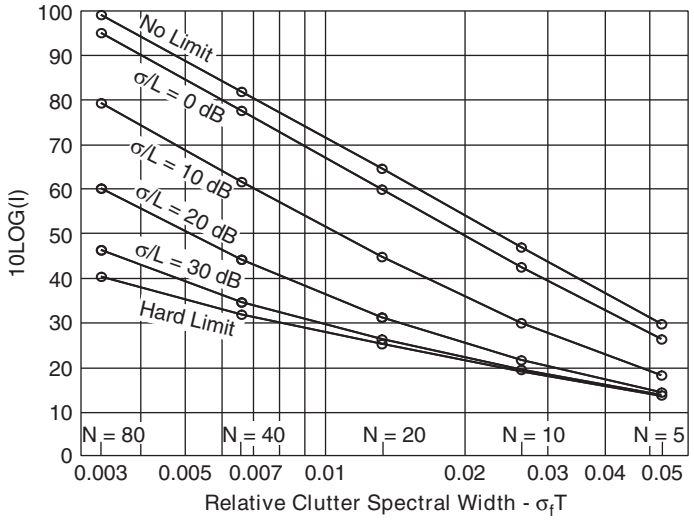


FIGURE 2.66 Mean improvement factor restriction versus amount of limiting and clutter spectral spread for a four-pulse canceler (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007 and H. R. Ward and W. W. Shrader*[34] *© IEEE 1968*)

Note that this discussion of binary detection is addressed to the spectral distribution of real clutter, that, when viewed in the time domain before limiting, has a smoothly varying change of the amplitude and phase of the clutter vector. This is distinct from clutter variations due to system instabilities that are noise-like, wherein the system dynamic range should be limited to prevent the instability residue from exceeding the system noise level.

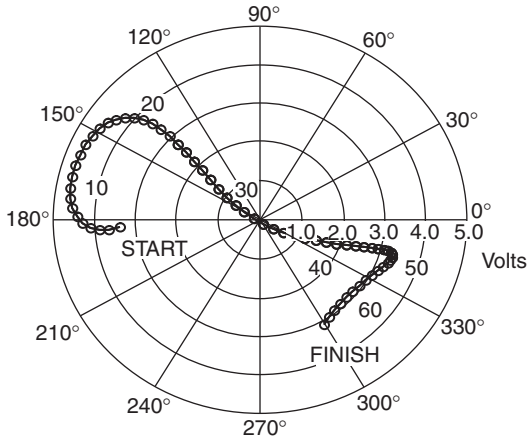


FIGURE 2.67 Polar representation of a linear clutter sequence for 20 hits per beamwidth (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007*)

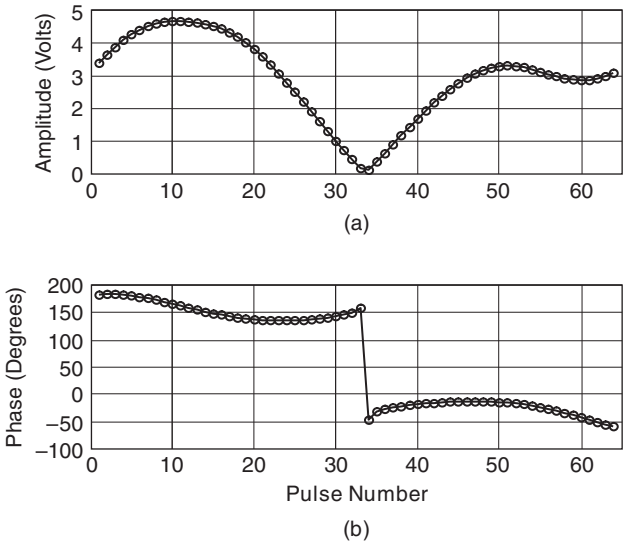


FIGURE 2.68 Linear clutter sequence amplitude and phase for 20 hits per beamwidth (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007*)

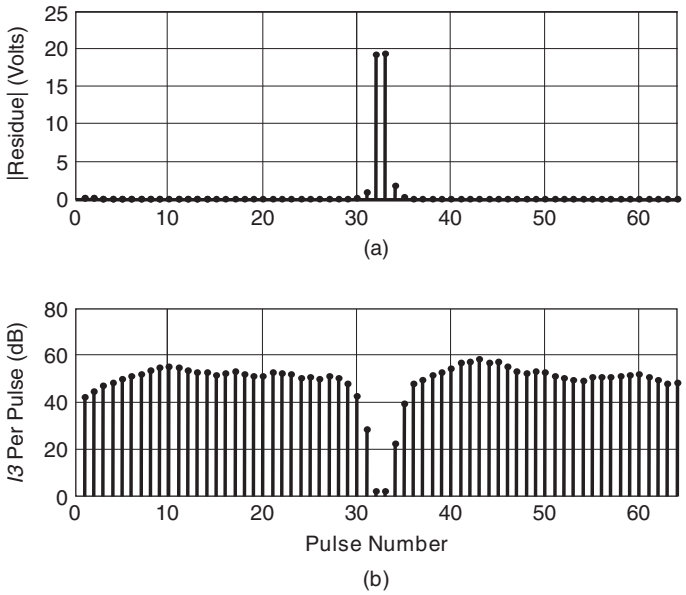


FIGURE 2.69 (*a*) Three-pulse canceler residue and (*b*) improvement factor for hard-limited clutter sequence for $N = 20$ hits per beamwidth (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007*)

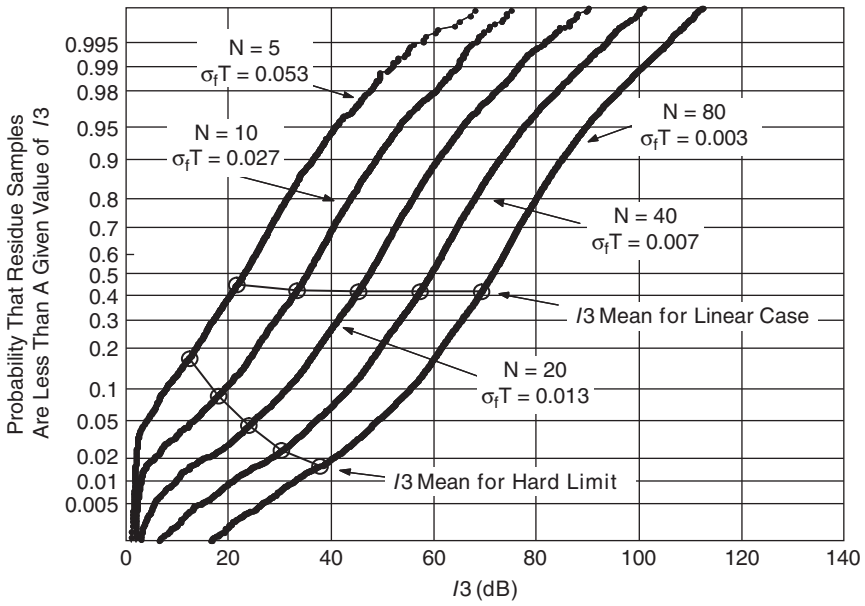


FIGURE 2.70 Distribution of I_3 and mean of I_3 for hard-limited clutter for different numers of scanning hits per beamwidth. For reference, the mean of I_3 is also shown for linear processing. (I_3 refers to the improvement factor of a three-pulse MTI canceler.) (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007*)

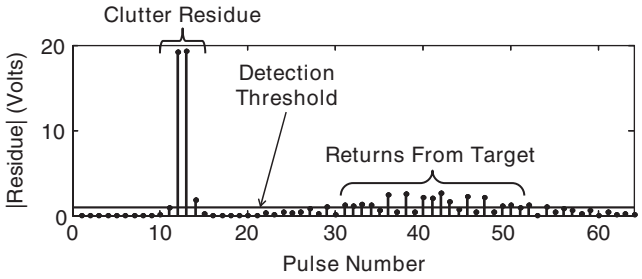


FIGURE 2.71 After MTI processing of the hard-limited distributed clutter sequence ($N = 20$) and a target superimposed on the clutter sequence, the residue spikes are distinctly different from the target returns. A binary M-of-N detector will reject the residue and keep the target. (*after T. M. Hall and W. W. Shrader*[32] *© IEEE 2007*)

2.12 RADAR SYSTEM STABILITY REQUIREMENTS

System Instabilities. Not only do the antenna motion and clutter spectrum affect the improvement factor that is attainable, but system instabilities also place a limit on MTI performance. These instabilities come from the stalo and coho, from the transmitter pulse-to-pulse frequency change if a pulsed oscillator and from pulse-to-pulse

phase change if a power amplifier, from the inability to lock the coho perfectly to the phase of the reference pulse, from time jitter and amplitude jitter on the pulses, and from quantization noise of the A/D converter.[36,37]

Phase instabilities will be considered first. If the phases of consecutive received pulses relative to the phase of the coho differ by, say, 0.01 rad, a limitation of 40 dB is imposed on I. The 0.01-rad clutter vector change would be equivalent to a target vector, 40 dB weaker than the clutter, being superimposed on the clutter, as shown in Figure 2.72.

In the power amplifier MTI system shown in Figure 2.73, pulse-to-pulse phase changes in the transmitted pulse can be introduced by the pulsed amplifier. The most common cause of a power amplifier introducing phase changes is ripple on the high-voltage power supply. Other causes of phase instability include ac voltage on a transmitter tube filament and uneven power supply loading, such as that caused by pulse-to-pulse stagger.

In the pulsed oscillator system, shown in Figure 2.74, pulse-to-pulse frequency changes result in phase run-out during the transmitted pulse. Phase run-out is the change of the transmitted pulse phase during the pulse duration with respect to the phase of the reference oscillator. If the coho locked perfectly to the end of the transmitted pulse, a total phase run-out of 0.02 rad during the transmitted pulse would then place an average limitation of 40 dB on the improvement factor attainable. Pulse-to-pulse frequency change in microwave oscillators is primarily caused by high-voltage power supply ripple. In the pulsed oscillator system, a pulse-to-pulse phase difference of 0.01 rad in locking the coho results in I limitation of 40 dB. (As noted elsewhere, frequency change during a pulse from a pulsed oscillator does not limit I if it repeats precisely pulse to pulse.)

The limitations on the improvement factor that are due to equipment instabilities in the form of frequency changes of the stalo and coho between consecutive transmitted pulses are a function of the range of the clutter. These changes are characterized in two ways. All oscillators have a noise spectrum. In addition, cavity oscillators, used because they are readily tunable, are microphonic, and thus their frequency may vary at an audio rate. The limitation on the improvement factor due to frequency changes is the difference in the number of radians that the oscillator runs through between the time of transmission and the time of reception of consecutive pulses. Thus, the improvement factor will be limited to 40 dB if $2\pi\Delta fT = 0.01$ rad, where Δf is the oscillator frequency change between transmitted pulses and T is the transit time of the pulse to and from the target.

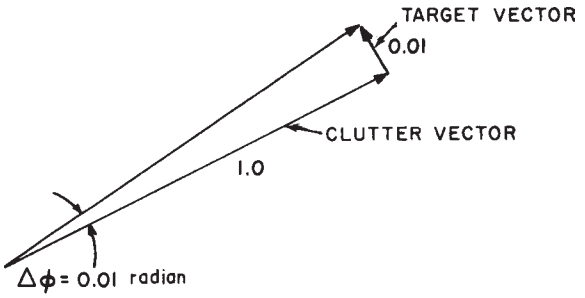


FIGURE 2.72 Phase instability

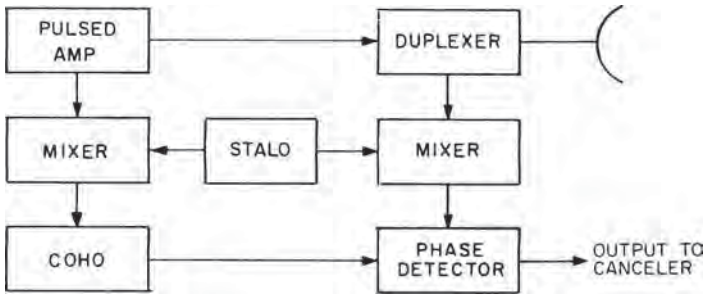


FIGURE 2.73 Power amplifier simplified block diagram

To evaluate the effects of oscillator phase noise on MTI performance, there are four steps. First, determine the single-sideband power spectral density of the phase noise as a function of frequency from the carrier.[38,39] Second, increase this spectral density by 6 dB. This accounts for a 3-dB increase because both sidebands of noise affect clutter residue, and a 3-dB increase because the oscillator contributes noise during both transmitting and receiving. Third, adjust the oscillator phase noise spectral density determined above due to the following three effects: (*a*) the self-cancellation of phase noise based on correlation resulting from the two-way range delay of the clutter of interest, (*b*) noise rejection due to the frequency response of the clutter filters, and (*c*) noise rejection due to the frequency response of the receiver passband. Finally, as the fourth step, integrate the adjusted spectral density of the phase noise across the entire passband. The result is the limitation on I due to the oscillator noise.

Rather than performing this integration of the residual noise numerically, a much simpler analysis can be carried out if both the oscillator phase noise characteristic and all of the adjustments to phase noise are approximated by straight lines on a decibel-versus-log frequency plot. This procedure becomes particularly simple when a MTI FIR filter using binomial coefficients is assumed. The locations along the frequency axis where the straight lines intersect are called *break frequencies.* This simplified procedure, which is similar to that presented in Vigneri et al.,[40] is described in the following paragraphs.

The first of the three adjustments—oscillator noise self-cancellation due to the range of the clutter of interest—reduces noise at the low frequencies by 20 dB per decade below the break frequency of $f = 1/(\sqrt{2} \cdot T_R \cdot \pi)$. Here $T_R = 2 \cdot R/c$ is the time

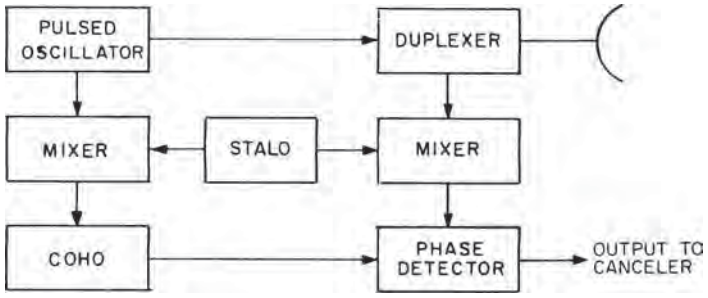


FIGURE 2.74 Pulsed oscillator simplified block diagram

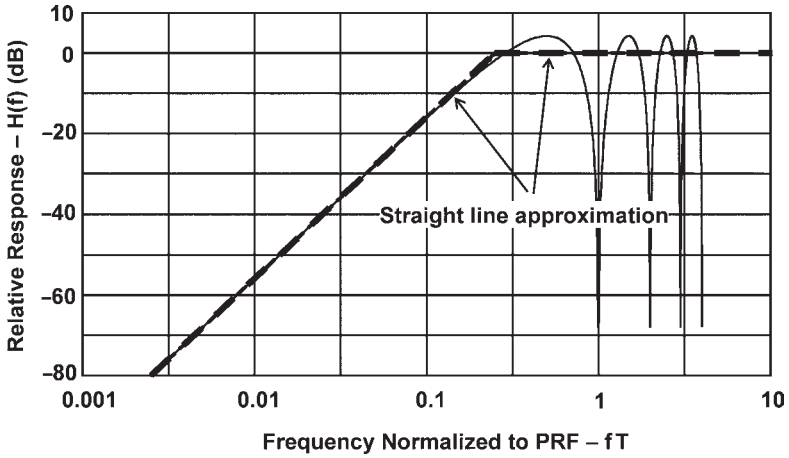


FIGURE 2.75 Straight-line approximation to two-delay binomial MTI

delay of the clutter return, R is the clutter range, and c is the speed of light. For the second adjustment due to the frequency response of the clutter filters, which as stated previously are assumed to be FIR cancelers with binomial weights, it is noted that the response at very low frequencies fall off at 20 dB per decade for one delay, 40 dB per decade for two delays, 60 dB per decade for three delays, etc. As an example, the approximation used for a two-delay MTI filter is shown in Figure 2.75. The MTI response has a peak value of $4/\sqrt{6} \approx 4.26$ dB, resulting in an average noise gain of unity, and the straight line approximation follows the low frequency asymptote up to the 0 dB level, which occurs at $fT = 0.249$, and stays constant at the 0 dB level at all higher frequencies. The justification for the 0 dB approximation at the higher frequencies is that the oscillator spectral density is more nearly constant and the average over one period of the MTI response is unity. For other binomial coefficient MTI cancelers, the break frequencies for the start of the response falloff are $fT = 0.225$ for one delay, 0.249 for two delays, 0.262 for three delays, and 0.271 for four delays.

For example, consider an oscillator with single-sideband phase-noise spectral density as shown in Figure 2.76. All oscillator noise contributions are assumed to be combined into this one curve. The single-sideband noise is increased by 3 dB because both sidebands affect system stability, and the power integration is only carried out for positive frequencies and by an additional 3 dB because the oscillator introduces noise in both the upconversion to the transmitted signal and in the receiver downconversion process.

Figure 2.77 shows the spectral modifications due to the system responses: (*a*) The first modification accounts for correlation due to the range to the clutter of interest [assumed clutter range is ≈ 50 nmi (92.6 km); thus, the break frequency is 365 Hz]. (*b*) Second, a three-pulse binomial-weighted canceler is assumed with the radar operating at a PRF of 360 Hz. Thus, the break frequency is $0.249 \times 360 = 90$ Hz. (*c*) Third, the receiver passband is assumed to extend from −500 kHz to +500 kHz with respect to the IF center frequency (1 MHZ total passband) at the −3 dB points and determined by a two-pole filter. Thus, the receiver passband response falls off at 40 dB per decade from the break frequency at 500 kHz as shown.

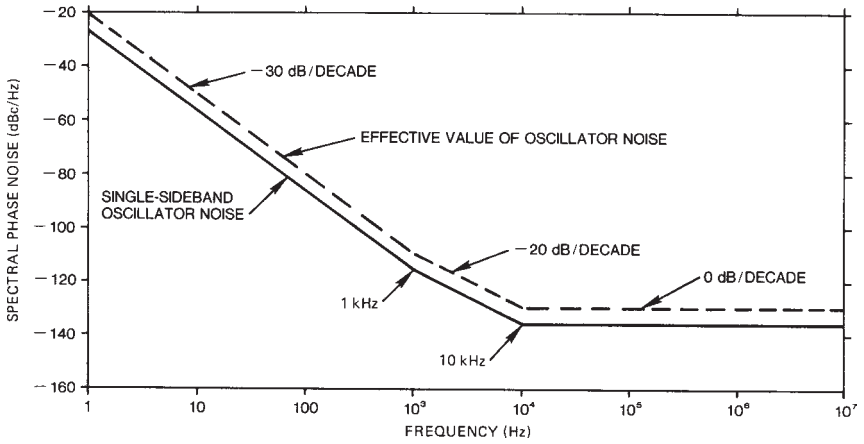


FIGURE 2.76 Single-sideband phase-noise spectral density of a microwave oscillator and the effective noise density

The adjusted phase-noise spectral density is shown in Figure 2.78. The total noise power with respect to the carrier is determined by integration of the noise power under the curve. The equation for the power spectral density of any one segment as a function of frequency is

$$S(f) = S_1 \cdot \left(\frac{f}{f_1}\right)^{\alpha} \qquad f_1 \le f \le f_2 \tag{2.42}$$

Here f_1 and f_2 are the start and end frequencies of the segment, respectively; S_1 (Hz^{-1}) is the phase noise spectral density relative to the carrier at the beginning of the segment and α is the slope of the segment in $\log_{10}$-units per decade. Note that the

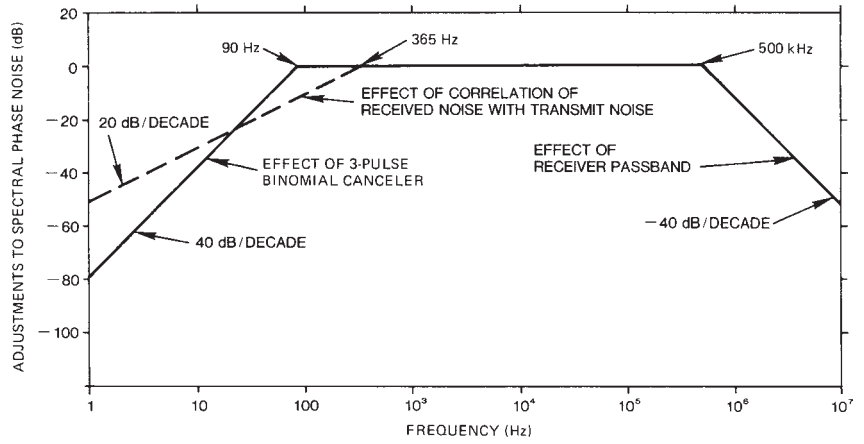


FIGURE 2.77 Adjustments, based on system parameters (see text), to the phase noise of a microwave oscillator

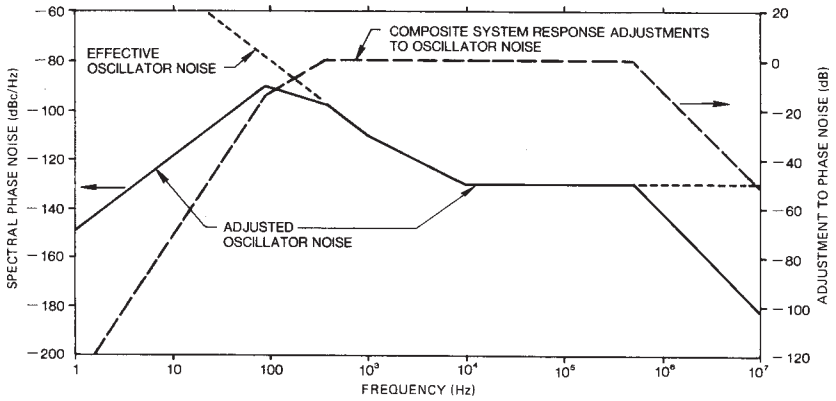


FIGURE 2.78 Composite adjustments and adjusted phase-noise spectral density

dBc/Hz values in Figure 2.78 correspond to $10 \cdot \log(S)$. Further denoting the phase noise spectral density relative to the carrier at the end of the segment as S_2 (Hz^{-1}), the slope is defined by

$$\alpha = \frac{\log_{10}(S_2/S_1)}{\log_{10}(f_2/f_1)} \tag{2.43}$$

The slope in dB/decade is equal to $10 \cdot \alpha$. The noise power contribution corresponding to this segment is found as

$$P = \begin{cases} \dfrac{S_1}{f_1^{\alpha}} \cdot \dfrac{1}{1+\alpha} \cdot \left[f_2^{\alpha+1} - f_1^{\alpha+1} \right] & all\ \alpha \neq -1 \\ \dfrac{S_1}{f_1^{\alpha}} \cdot [\ln(f_2) - \ln(f_1)] & \alpha = -1 \end{cases} \tag{2.44}$$

Table 2.4 gives the integration for the example. When the integrated powers for all segments have been calculated, they are summed and then converted back to dBc. The final answer, –66.37 dBc, is the limit on I that results from oscillator noise. The limit on I_{SCR} (dB) is I (dB) plus target integration gain (dB).

TABLE 2.4 Integration of the Phase-Noise Spectral Density of Figure 2.76 with Adjustments of Figure 2.77 as Shown in Figure 2.78

Segment	f_1, Hz	f_2, Hz	Slope, dB/dek	Slope α	S_1 dBc/Hz	S_2 dBc/Hz	Integrated power	Integrated power, dBc
1	1	90	30.0	3	–149.4	–90.8	0.188e-7	–77.25
2	90	365	–10.0	–1	–90.8	–96.9	0.105e-6	–69.80
3	365	1,000	–30.0	–3	–96.9	–110.0	0.323e-7	–74.91
4	1.0e3	1.0e4	–20.0	–2	–110.0	–130.0	0.900e-8	–80.46
5	1.0e4	5.0e5	0.0	0	–130.0	–130.0	0.490e-7	–73.10
6	5.0e5	1.0e7	–40.0	–4	–130.0	–182.0	0.167e-7	–77.78
				Total integrated noise power			0.231e-6	–66.37

Time jitter of the transmitted pulses results in degradation of MTI systems. Time jitter results in failure of the leading and trailing edges of the pulses to cancel, the amplitude of each uncancelled part being $\Delta t/\tau$, where Δt is the time jitter and τ is the transmitted pulse length. The total residue power is $2(\Delta t/\tau)^2$, and therefore the limitation on the improvement factor due to time jitter is $I = 20 \cdot \log[\tau/(\sqrt{2}\Delta t)]$ (dB). This limit on the improvement factor is based on a CW transmitter pulse and on the assumption that the receiver bandwidth is matched to the duration of the transmitted pulse. In a pulse compression system, the receiver bandwidth is wider by the time-bandwidth ($B\tau$) product; thus the clutter residue power at each end of the pulse increases in proportion to the $B\tau$ product. The limit on I for a chirp pulse compression system is then $I = 20 \cdot \log[\tau/(\sqrt{2} \cdot \Delta t \cdot \sqrt{B \cdot \tau})]$. For pulse compression systems employing phase-coded waveforms, the factor 2 in the preceding equation should be multiplied by the number of subpulses in the waveform. Thus, for example, the limit on I for a 13-pulse Barker code is

$$I = 20 \log \ [\tau/(\sqrt{2 \times 13} \ \ \Delta t\sqrt{13})] \text{dB} \tag{2.45}$$

Pulse-width jitter results in one-half the residue of time jitter, and

$$I = 20 \log \ \frac{\tau}{\Delta \text{PW}\sqrt{B\tau}} \text{dB} \tag{2.46}$$

where ΔPW is pulse-width jitter.

Amplitude jitter in the transmitted pulse also causes a limitation of

$$I = 20 \log \ \frac{A}{\Delta A} \text{dB} \tag{2.47}$$

where A is the pulse amplitude and ΔA is the pulse-to-pulse change in amplitude. This limitation applies even though the system uses limiting before the canceler because there is always much clutter present that does not reach the limit level. With most transmitters, however, the amplitude jitter is insignificant after the frequency-stability or phase-stability requirements have been met.

Jitter in the sampling time in the A/D converter also limits MTI performance. If pulse compression is done prior to the A/D or if there is no pulse compression, this limit is

$$I = 20 \log \ \frac{\tau}{J\sqrt{B\tau}} \text{dB} \tag{2.48}$$

where J is the timing jitter, τ is transmitted pulse length, and $B\tau$ is the time-bandwidth product. If pulse compression is done subsequent to the A/D converter, then the limitation is

$$I = 20 \log \ \frac{\tau}{JB\tau} \text{dB} \tag{2.49}$$

The limitations on the attainable MTI improvement factor are summarized in Table 2.5. This discussion has assumed that the peak-to-peak values of these instabilities occur on a pulse-to-pulse basis, which is often the case in pulse-to-pulse staggered MTI operation. If it is known that the instabilities are random, the peak

TABLE 2.5 Instability Limitations

Pulse-to-Pulse Instability	Limit on Improvement Factor
Oscillator phase noise	*See discussion in text.*
Transmitter frequency	$I = 20 \log [1/(\pi \Delta f \tau)]$
Stalo or coho frequency	$I = 20 \log [1/(2\pi \Delta f T)]$
Transmitter phase shift	$I = 20 \log (1/\Delta\phi)$
Coho locking	$I = 20 \log (1/\Delta\phi)$
Pulse timing	$I = 20 \log [\tau/(\sqrt{2}\Delta t\sqrt{B\tau})]$
Pulse width	$I = 20\log [\tau/(\Delta PW\sqrt{B\tau})]$
Pulse amplitude	$I = 20 \log (A/\Delta A)$
A/D jitter	$I = 20 \log [\tau(J\sqrt{B\tau})]$
A/D jitter with pulse compression following A/D	$I = 20 \log [\tau/(JB\tau)]$

where Δf interpulse frequency change
τ transmitted pulse length
T transmission time to and from target
$\Delta\phi$ interpulse phase change
Δt time jitter
J A/D sampling time jitter
$B\tau$ time-bandwidth product of pulse compression system ($B\tau$ = unity for CW pulses)
ΔPW pulse-width jitter
A pulse amplitude, V
ΔA interpulse amplitude change

values shown in these equations can be replaced by the rms pulse-to-pulse values, which gives results essentially identical to Steinberg's results.[41]

If the instabilities occur at some known frequency, e.g., high-voltage power supply ripple, the relative effect of the instability can be determined by locating the response on the velocity response curve for the MTI system for a target at an equivalent doppler frequency. If, for instance, the response is 6 dB down from the maximum response, the limitation on I is about 6 dB less severe than indicated in the equations in Table 2.5. If all sources of instability are independent, as would usually be the case, their individual power residues can be added to determine the total limitation on MTI performance.

Intrapulse frequency or phase variations do not interfere with good MTI operation provided they repeat precisely from pulse to pulse. The only concern is a loss of sensitivity if phase run-out during the transmitted pulse or mistuning of the coho or stalo permits the received pulses to be significantly detuned from the intended IF frequency. If a 1-rad phase run-out during the pulse is permitted, the system detuning may be as large as $1/(2\pi\tau)$ Hz with no degradation of MTI performance.

To give an example of interpulse stability requirements, consider a 3000-MHz radar transmitting a CW pulse of duration $\tau = 2$ μs and the requirement that no single system instability will limit the MTI improvement factor attainable at a range of 100 nmi to less than 50 dB, a voltage ratio of 316:1. The rms pulse-to-pulse transmitter frequency change (if a pulsed oscillator) must be less than

$$\Delta f = \frac{1}{316\pi\tau} = 504 \text{ Hz} \tag{2.50}$$

which is a stability of about 2 parts in 10^7.

The rms pulse-to-pulse transmitter phase-shift change (if a power amplifier) must be less than

$$\Delta\phi = \frac{1}{316} = 0.00316 \text{ rad} = 0.18^0 \tag{2.51}$$

The stalo or coho frequency change in the interpulse period must be less than

$$\Delta f = \frac{1}{316\,(2\pi)\;\;(100 \times 12.36 \times 10^{-6})} = 0.4 \;\; \text{Hz} \tag{2.52}$$

which is a stability of 1 part in 10^{10} for the stalo (at about 3 GHz) and 1 part in 10^8 for the coho (assuming a 30-MHz IF frequency).

The coho locking (if a pulsed oscillator system) must be within

$$\Delta\phi = \frac{1}{316} = 0.00316 \text{ rad} = 0.18^0 \tag{2.53}$$

The pulse timing jitter must be less than

$$\Delta t = \frac{\tau}{316\sqrt{2}\sqrt{1}} = \frac{2 \times 10^{-6}}{316\sqrt{2}} = 4.5 \times 10^{-9} \text{ s} \tag{2.54}$$

The pulse-width jitter must be less than

$$\Delta\text{PW} = \frac{\tau}{316\sqrt{1}} = \frac{2 \times 10^{-6}}{316} = 6 \times 10^{-9} \text{ s} \tag{2.55}$$

The pulse amplitude change must be less than

$$\frac{\Delta A}{A} = \frac{1}{316} = 0.00316 = 0.3 \text{ percent} \tag{2.56}$$

The A/D sampling time jitter must be less than

$$J = \frac{\tau}{316\sqrt{1}} = \frac{2 \times 10^{-6}}{316} = 6 \times 10^{-9} \text{ s} \tag{2.57}$$

Of the above requirements, oscillator phase noise may dominate. However, in systems with large bandwidths (short compressed pulses), the timing jitter requirements become significant and may require special clock regeneration circuitry at key system locations.

Effect of Quantization Noise on Improvement Factor. Quantization noise, introduced in the A/D converter, limits the attainable MTI improvement factor. Consider a conventional video MTI system, as shown in Figure 2.79. Because the peak signal level is controlled by the linear-limiting amplifier, the peak excursion of the phase-detector output is known, and the A/D converter is designed to cover this excursion. If the A/D converter uses N bits and the phase-detector output is from −1 to +1, the quantization interval is $2/(2^N - 1)$. The rms value of the signal-level deviation

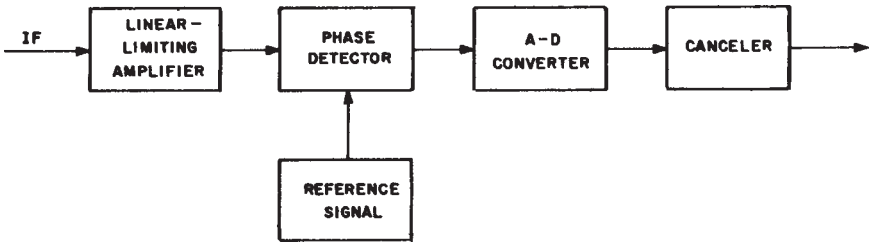


FIGURE 2.79 Digital MTI consideration

introduced by the A/D converter is $2/[(2^N - 1)\sqrt{12}]$. The limit on the MTI improvement factor that this imposes on a signal reaching the full excursion of the phase detector is found by substituting in the following equation from Table 2.5:

$$I = 20 \log \frac{A}{\Delta A} = 20 \log \left\{ \frac{1}{[(2^N - 1)\sqrt{3.0}]^{-1}} \right\} = 20 \log [(2^N - 1)\sqrt{3.0}] \tag{2.58}$$

Because two quadrature channels contribute independent A/D noise, the average limit on the improvement factor of a full-range signal is

$$I = 20 \log \left[(2^N - 1)\sqrt{\frac{3.0}{2}} \right] = 20 \log [(2^N - 1)\sqrt{1.5}] \tag{2.59}$$

If the signal does not reach the full excursion of the A/D converter, which is normally the case, then the quantization limit on I is proportionately more severe. For example, if the system is designed so that the mean level of the strongest clutter of interest is 3 dB below the A/D converter peak, the limit on I would be $20 \cdot \log_{10}[(2^N - 1) \cdot \sqrt{0.75}]$. (This is tabulated in Table 2.6.)

This discussion of A/D quantization noise has assumed perfect A/D converters. Many A/D converters, particularly under high-slew-rate conditions, are less than perfect. This, in turn, leads to system limitations more severe than predicted here (see Section 2.13).

TABLE 2.6 Typical Limitation on I Due to A/D Quantization

Number of Bits, N	Limit on MTI Improvement Factor I, dB
10	59.0
11	65.0
12	71.0
13	77.0
14	83.0
15	89.1
16	95.1

Pulse Compression Considerations.† When an MTI system is used with pulse compression, the system target detection capability in clutter may be as good as a system transmitting the equivalent short pulse, or the performance may be no better than a system transmitting the same length CW pulse. The kind of clutter environment, the system instabilities, and the signal processing utilized determine where the system performance will fall between the above two extremes. Unless provision is incorporated for coping with system instabilities and clutter spectral spread, the MTI pulse compression system may fail to work at all in a clutter environment.

Ideally, a pulse compression receiver coupled with an MTI would appear as in Figure 2.80*a*.‡ If the pulse compression system was perfect, the compressed pulse would look as if the radar had transmitted and received a short pulse, and MTI processing could proceed as if the pulse compression had not existed. In practice, the compressed pulse will have time sidelobes from three basic causes. The first is waveform and system design, which includes components that may be nonlinear with frequency, etc. These sidelobes will be stable. That is, they should repeat precisely on a pulse-to-pulse basis and thus will cancel in the MTI canceler. It is assumed that the radar system is fully coherent as required by rule 3 in Section 2.17. The second cause of pulse compression sidelobes is system instabilities, such as noise on local oscillators, transmitter time jitter, transmitter tube noise, and A/D converter jitter. These sidelobes are noise-like and are proportional to the clutter amplitude. They will not cancel in the MTI canceler. The third source of sidelobes is high-frequency ripple in the transmitter power supply.

If the transmitter power supply incorporates high-frequency ac-dc and/or dc-dc converters, and if the converter frequency components are not sufficiently filtered, there will be discrete time sidelobes, offset from the clutter in range, as predicted by paired-echo theory.[42] The paired-echo sidelobes will also have a doppler frequency equal to the converter frequency. This frequency (f_{conv}) will alias into the PRF (f_r) doppler interval at the frequency (f_{dop}) [f_{dop} = modulo (f_{conv}, f_r)]). These sidelobes will not cancel unless the high-frequency converters are synchronized to a multiple of the PRF, in which case $f_{\text{dop}} = 0$.

Assume that the noise-like component of the sidelobes is down 50 dB from the peak transmitted signals. This noise-like component will not cancel in the MTI system, and therefore, for each clutter area that exceeds the system threshold by 50 dB or more, the residue will exceed the detection threshold. If the clutter exceeds the threshold by 70 dB, the residue from the MTI system will exceed the detection threshold by 20 dB, eliminating the effectiveness of the MTI. Figure 2.80*b* shows a sketch of this effect.

To ensure that the noise-like pulse-compression sidelobes will not exceed the system noise after the MTI canceler, the system stability budget must ensure that the instability sidelobe level is lower than the dynamic range of the receiving system. The receiving system dynamic range is ultimately determined (in a well-designed system) by the IF

† All signal processing following the A/D detector is done digitally. It is more meaningful, however, to describe and depict the processing in an analog manner.

‡ The IF bandpass limiter [*Radar Handbook, 2nd Ed.*, pp. 3.30–3.32] shown in this and subsequent diagrams has an amplitude output characteristic that is linear for input signal voltages from noise level to within 6 dB of the limiter output maximum voltage and then transitions smoothly to the maximum output voltage.[32] The phase of the input signal is precisely preserved. These limiter characteristics exist whether the filter is implemented in analog circuitry or a digital algorithm.

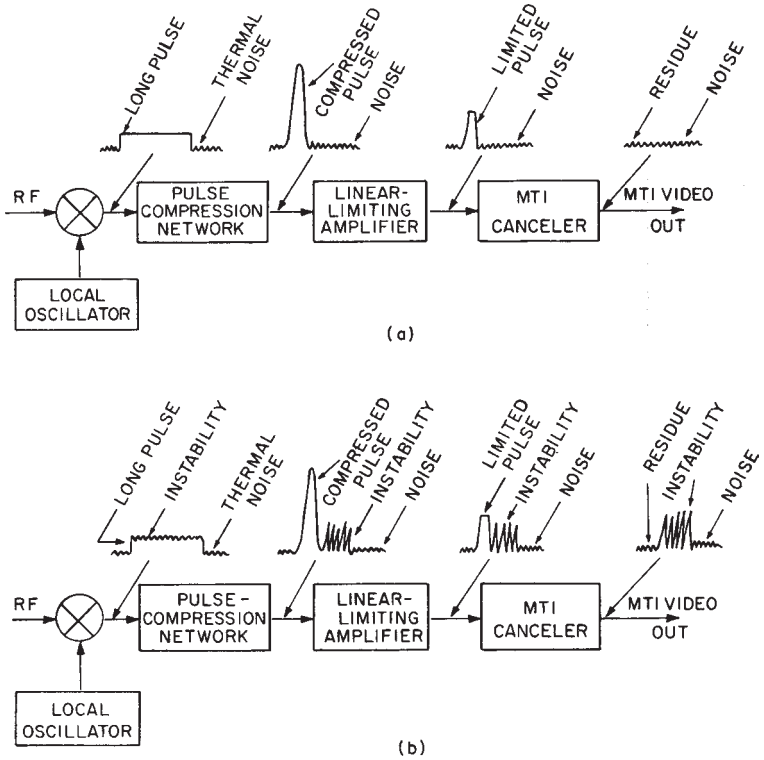


FIGURE 2.80 Pulse compression with MTI: (*a*) ideal but difficult-to-achieve combination and (*b*) effect of oscillator on transmitter instabilities

bandpass limiter that precedes the A/D converter. If system instabilities cannot be controlled to be less than the system dynamic range, then the system dynamic range should be decreased. (An alternative to decreasing the dynamic range is to depend on a cell-averaging constant false alarm rate (CA-CFAR) processor after the signal processing to provide a threshold that rides over the residue noise, but the efficacy of this method depends on the residue noise being completely noise-like, which is unlikely.)

After addressing the unstable pulse-compression sidelobes, it is still necessary to control detections from residue caused by the spectral spread of the clutter or by low frequency transmitter power supply ripple. This can be accomplished by limiting the maximum signal amplitude at the input to the canceler. The process described above is depicted in Figure 2.81.

One approach that has been successful in achieving the maximum MTI system performance attainable within the limits imposed by system and clutter instabilities

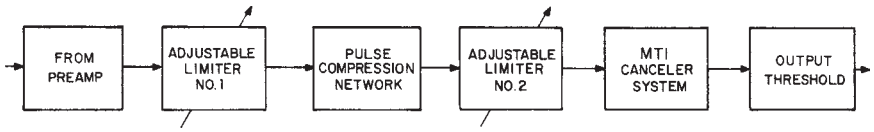


FIGURE 2.81 Practical MTI pulse-compression combination

is shown in Figure 2.81. (Transmitter noise will be used in the following discussion to represent all possible system instabilities that create noise-like pulse-compression time sidelobes.)

Limiter 1 is set to limit the system dynamic range to the range between peak clutter and clutter instability noise. Limiter 2 is set so that the dynamic range at its output is equal to the expected MTI improvement factor as limited by clutter spectral spread or low-frequency transmitter power supply ripple. These limiter settings cause the residue due to transmitter noise and the residue due to other instabilities, such as quantization noise and internal-clutter motion, each to be equal to front-end thermal noise at the canceler output. This allows maximum sensitivity without an excessive false-alarm rate. Limiter 1 is a very efficient constant-false-alarm-rate device against system instabilities because it suppresses the instability noise in direct proportion to the clutter signal strength but does not suppress at any time when the clutter signal is not strong. Although the limiters cause partial or complete suppression of some desired targets in the clutter areas, no targets are suppressed that could otherwise have been detected in the presence of clutter residue at the system output if the limiters had not been used.

As a specific example, consider a system with a pulse-compression ratio of about 30 dB and system instability noise approximately 28 dB below the carrier power. Assume that the MTI canceler improvement factor is 30 dB, limited by clutter spectral spread. With the above system parameters, a receiver system that will provide the maximum obtainable performance is shown in Figure 2.82. At the output of the pulse-compression network, the system instability noise will be equal to or less than thermal noise for either distributed clutter or point clutter, and the peak clutter signals will vary from about 28 dB above thermal noise for evenly distributed clutter to 58 dB above thermal noise for strong point clutter.

Because the MTI canceler is expected to attenuate clutter by 30 dB, the second limiter is provided to prevent the residue from strong clutter from exceeding the threshold. Without the second limiter, a strong-point reflector that was 58 dB above noise at the canceler input would have a residue 28 dB above noise at the canceler output. This would be indistinguishable from an aircraft target.

It the transmitter noise were 15 dB less than assumed above, the first limiter would be set 43 dB above thermal noise and much less target suppression would occur. Thus target detectability would improve in and near the strong clutter areas even though the MTI improvement factor was still limited to 30 dB by internal-clutter motion.

In summary, the noise-like pulse compression sidelobes and the duration of the uncompressed pulse dictate how effective a pulse-compression MTI system can be. Systems have been built in which transmitter noise and long uncompressed pulses combined to make the systems incapable of detecting aircraft targets in or near land clutter. Some existing pulse-compression systems have not deliberately provided the

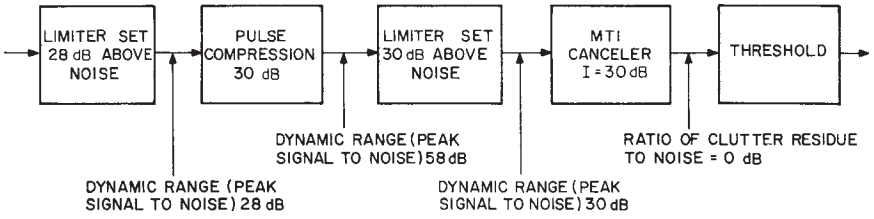

LIMITER SET 28 dB ABOVE NOISE
PULSE COMPRESSION 30 dB
LIMITER SET 30 dB ABOVE NOISE
MTI CANCELER I = 30 dB
THRESHOLD
DYNAMIC RANGE (PEAK SIGNAL TO NOISE) 58 dB
RATIO OF CLUTTER RESIDUE TO NOISE = 0 dB
DYNAMIC RANGE (PEAK SIGNAL TO NOISE) 28 dB
DYNAMIC RANGE (PEAK SIGNAL TO NOISE) 30 dB

FIGURE 2.82 MTI with pulse compression

two separate limiters described above, but the systems work because dynamic range is sufficiently restricted by circuit components. Other systems, such as those that deliberately hard-limit before pulse compression for CFAR reasons, do not have clutter residue problems but suffer from significant target suppression in the clutter areas.

An alternative to the use of limiters is the use of clutter maps in conjunction with the CA-CFAR. Clutter maps work well for stationary radars operating at fixed frequencies, but are less effective for other radars. The CA-CFAR is useful, even for a system with IF limiters, because there will be small variations (on the order of a few dB) in the combination of clutter residue and system noise. To reemphasize, however, without the limiters, there may be tens of dB's difference between clutter residue and system noise.

2.13 DYNAMIC RANGE AND A/D CONVERSION CONSIDERATIONS

The accurate conversion of the radar IF signal into a digital representation of the complex envelope is an important step in the implementation of a modern digital signal processor. This analog-to-digital (A/D) conversion must preserve the linearity of amplitude and phase over the required dynamic range, have a small effect on overall radar system noise temperature, and be free from undesired spurious responses.

Advances in A/D converter technology is now making it possible to directly convert an analog IF signal into a corresponding digital complex representation, rather than going through the intermediate step of first downconverting the IF signal into baseband in-phase (I) and quadrature (Q) components and subsequently using a separate A/D converter in each of these two channels.

A flow chart of a direct IF A/D converter is illustrated in Figure 2.83 along with spectral representations of the signal throughout the conversion process. The IF input centered at the frequency f_{IF} is first passed through a bandpass filter to ensure that negligible aliasing will occur during the subsequent A/D conversion process. On the right in Figure 2.83, the top graph shows the positive and negative parts of the signal spectrum at the IF filter output. The positive part of this spectrum corresponds to the complex envelope, which needs to be translated into the digital I and Q representation. This filter output becomes the input to the A/D converter operating at a sampling rate of f_{AD}. The spectrum of the A/D converter output is again shown, and it is obtained simply by replicating the original IF spectrum from minus infinity to plus infinity with a period of f_{AD}. In this example, an A/D conversion rate of $f_{AD} = \frac{4}{3} \cdot f_{IF}$ is assumed. The optimum choice of the A/D converter sampling rate ensures that the negative part of the spectrum has the smallest possible overlap with the positive part of the spectrum.

The smallest possible overlap occurs when the A/D sampling rate is related to the radar IF frequency as follows[43]:

$$f_{AD} = \frac{4 \cdot f_{IF}}{2 \cdot M - 1} \tag{2.60}$$

where M is an integer greater than 1. Thus, optimum sampling rates are $4\,f_{IF}$, $1.3333\,f_{IF}$, $0.8\,f_{IF}$, $0.57\,f_{IF}$,… etc. The corresponding maximum unaliased (or Nyquist) bandwidth is $B_{NQ} = f_{AD}/2$. This value is, therefore, the maximum allowable cutoff bandwidth of the IF bandpass filter at the input to the A/D converter. It is not strictly necessary to use

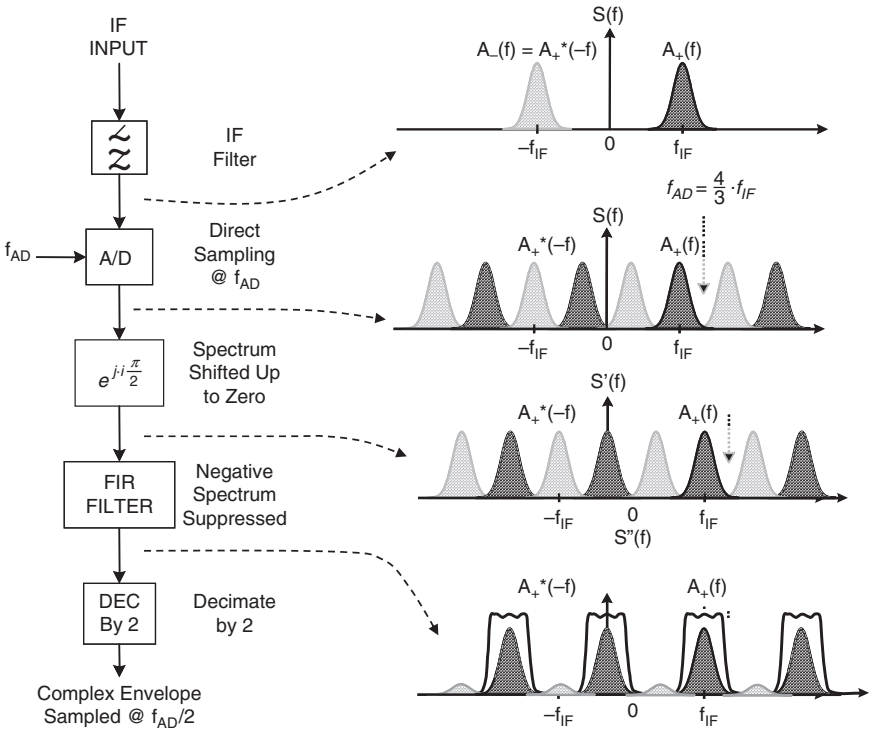


FIGURE 2.83 Implementation of A/D conversion using direct sampling of the IF signal

an A/D converter sampling rate as given by Eq. 2.60, but other values will result in an available Nyquist bandwidth less than $f_{AD}/2$. This is shown in Figure 2.84 where the normalized Nyquist bandwidth is shown as a function of the relative A/D converter sampling rate. From this figure, it is seen that the direct conversion approach will fail whenever a value of *M,* which is located halfway between the optimum values, is used.

At the A/D converter output, the signal samples are still real valued. To be able to extract the complex envelope corresponding to the positive part of the spectrum, $2 \cdot A_+(f - f_{IF})$, it is necessary to shift the spectrum at the A/D converter output down in frequency by the amount f_{IF}. This corresponds to a multiplication by the time series $u(i) = e^{-j \cdot i \cdot \frac{3 \cdot \pi}{2}}$. Equivalently, the complex envelope spectrum below zero frequency can be shifted up to zero frequency by multiplication with the time series $u(i) = e^{j \cdot i \cdot \frac{\pi}{2}}$.This results in the spectrum shown where the desired spectrum corresponding to the complex envelope is centered at zero frequency, but the signal still contains the unwanted negative spectral components (light shading). As a result of this frequency translation, the signal has now become complex. A digital FIR bandpass filter with a nearly rectangular response is then applied to reject the negative-frequency components as shown in the final graph on the right. The desired sampled complex envelope representation has now been realized, but at the original sampling rate of f_{AD}. If desired, the oversampling can finally be removed through decimation by a factor of 2 as shown in the last step in the figure.

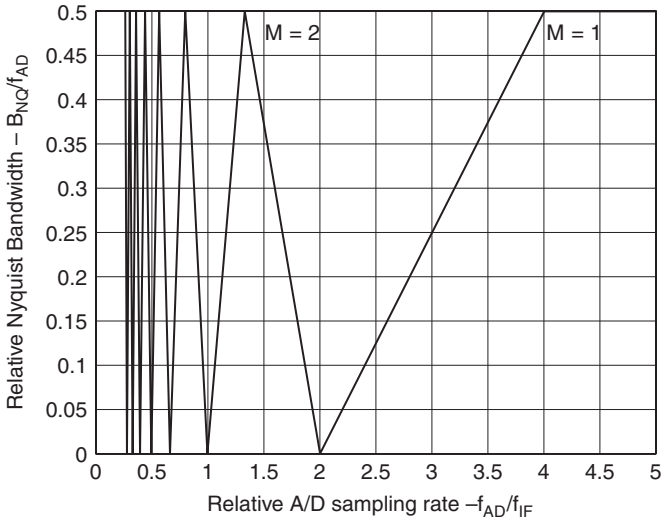


FIGURE 2.84 Available Nyquist bandwidth vs. A/D converter sampling rate

A/D converters are typically characterized by their signal-to-noise ratio (SNR) performance referred to a bandwidth equal to the A/D sampling rate. Often this SNR is not as high as one would expect based on the number of bits used by the A/D converter. Sometimes the actual performance of an A/D converter is characterized by an effective number of bits, smaller than the actual number of bits and corresponding to the achievable SNR. The SNR of an A/D converter sets an upper limit on the achievable improvement factor.

2.14 ADAPTIVE MTI

When the doppler frequency of the returns from clutter is unknown at the radar input, special techniques are required to guarantee satisfactory clutter suppression. As discussed in Section 2.10, the doppler filter bank will usually be effective against moving clutter. This requires that the individual filters be designed with a low sidelobe level in the regions where clutter may appear and that each filter be followed by appropriate CFAR processing circuits to reject unwanted clutter residue. When clutter suppression is to be implemented with a single MTI filter, it is necessary to use adaptive techniques to ensure that the clutter falls in the MTI rejection notch. An example of such an adaptive MTI is TACCAR,[31] originally developed for airborne radars. In many applications, the adaptive MTI will further have to take into account the situation where multiple clutter sources with different radial velocities are present at the same range and bearing.

Usually the doppler shift of clutter returns is caused by the wind field, and early attempts of compensating in the MTI have varied the coho frequency sinusoidally as a function of azimuth based on the average wind speed and direction. This approach is

unsatisfactory because the wind field rarely is homogeneous over a large geographical area and because the wind velocity usually is a function of altitude due to wind shear (important for rain clutter and chaff). Against a single clutter source, an implementation is required that permits the MTI clutter notch to be shifted as a function of range. An example of such an adaptive MTI implementation is shown in Figure 2.85. The phase-error circuit compares the clutter return from one sweep to the next. Through a closed loop, which includes a smoothing time constant, the error signal controls a phase shifter at the coho output such that the doppler shift from pulse to pulse is removed. It should be noted that since the first sweep entering the MTI is taken as a reference, any phase shift run-out as a function of range will increase proportionally to the number of sweeps. Ultimately this run-out will exceed the speed of response of the closed loop, and the MTI must be reset. This type of closed-loop adaptive MTI must, therefore, be operated for a finite set (batch) of pulses to ensure that this will not happen. Such batch-mode operation is also required if a combination of MTI operation and frequency agility is desired.

If a bimodal clutter situation is caused by the simultaneous presence of returns from land clutter and weather or chaff, an adaptive MTI can be implemented following a fixed-clutter-notch MTI section, as illustrated in Figure 2.86. The number of zeros used in the fixed- (zero doppler) clutter-notch section of the MTI is determined by the required improvement factor and the spectral spread of the land clutter. Typically, the fixed-notch MTI would use two or three zeros. For the adaptive portion of the MTI, a fully digital implementation is shown in which the pulse-to-pulse phase shift of the clutter output from the first canceler is measured and averaged over a given number of range cells. This estimated phase shift is added to the phase shift, which is applied to the data on the previous sweep, and this new phase shift is applied to the current data.

The range averaging must be performed separately on the I and Q components of the measured phase in each range cell due to the 2π ambiguity of the phase representation itself. The accumulation of the applied phase shift from sweep to sweep, however, must be performed directly on the phase and is computed modulo 2π. The number of zeros of the adaptive MTI section is again determined by the required improvement factor and the expected spectral spread of the clutter. The phase shift is applied to the input data in the form of a complex multiply, which again requires the transformation of the phase angle into rectangular coordinates. This transformation can easily be performed by a table lookup operation in a read-only memory.

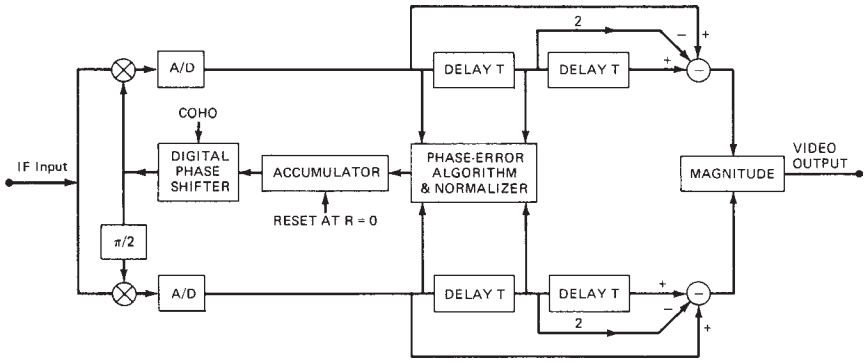


FIGURE 2.85 Block diagram of closed-loop adaptive digital MTI

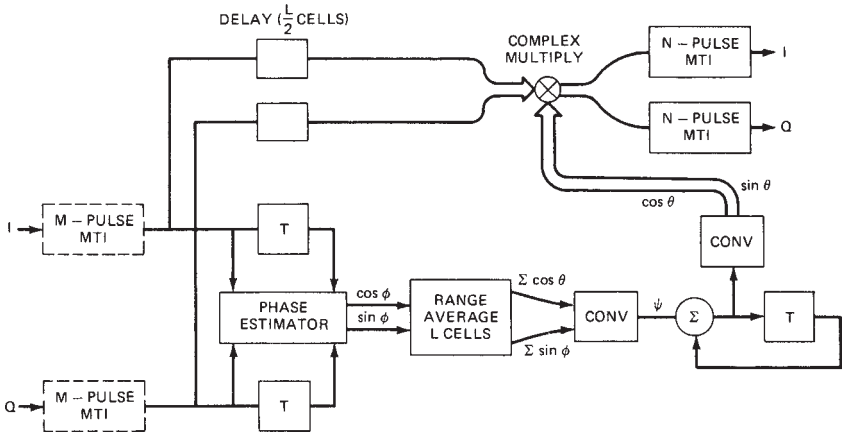


FIGURE 2.86 Open-loop adaptive MTI for cancellation of simultaneous fixed and moving clutter

When doppler shifts are introduced by digital means as described above, the accuracy of the I and Q representation of the original input data becomes an important consideration. Any dc offset, amplitude imbalance, quadrature phase error, or nonlinearity will result in the generation of undesired sidebands that will appear as residue at the canceler output. A discussion of A/D conversion considerations was presented in Section 2.13.

In the adaptive MTI implementation described above, the number of zeros allocated to each of the two cancelers was fixed, based on an a priori assessment of the clutter suppression requirement. The only variation possible would be to completely bypass one (or both) of the MTI cancelers if no land clutter or weather or chaff returns are received on a given radial. A more capable system can be implemented if the number of zeros can be allocated dynamically to either clutter source as a function of range. This leads to a fully adaptive MTI implementation using a more complex adaptation algorithm, as discussed below. Such an adaptive MTI may provide a performance close to the optimum discussed in Section 2.7.

In order to illustrate the difference in performance between such candidate MTI implementations, a specific example is considered next. For this example, land clutter returns are present at zero doppler with a normalized spectral spread of $\sigma_f T = 0.01$, and chaff returns are present at a normalized doppler offset of $f_d T = 0.25$ with a normalized spectral spread of $\sigma_f T = 0.05$. The power ratio of the land clutter to that of the chaff is denoted Q (dB). Thermal noise is not considered in this example. In both cases, the total number of filter zeros is assumed to be equal to 3. For the adaptive MTI with a fixed allocation of zeros, two zeros are located at zero doppler and the remaining zero is centered on the chaff returns. In the optimum MTI, the zero locations are chosen so that that overall improvement factor is maximized. The results of this comparison are presented in Figure 2.87, which shows the improvement factor for the optimum and the adaptive MTI as a function of the power ratio Q (dB). When Q is small so that chaff returns dominate, a significant performance improvement can be realized by using all MTI filter zeros to cancel the chaff returns. The performance difference for large values of Q is a result of an assumption made that the location of the third zero remains fixed at the chaff doppler frequency. In reality, the adaptive MTI would move

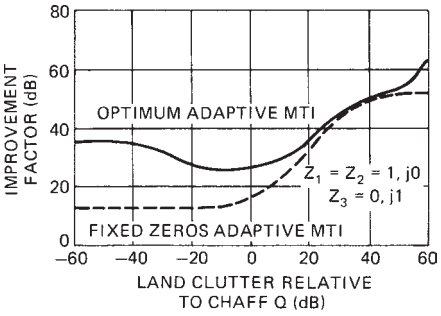


FIGURE 2.87 Improvement factor comparison of optimum and adaptive MTI against fixed and moving clutter of ratio Q

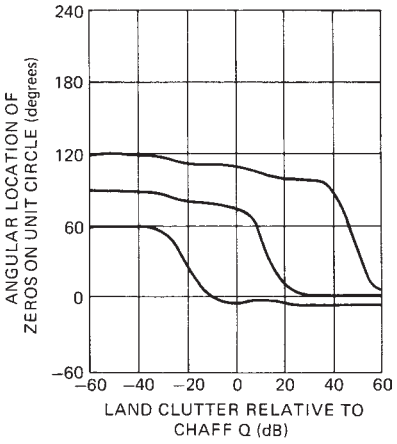


FIGURE 2.88 Location of the three filter zeros for an optimum MTI used against fixed and moving clutter

its third zero to the land clutter as the land clutter residue starts to dominate the output of the first canceler. The zero locations of the optimum MTI are shown in Figure 2.88 and can be seen to move between the land clutter at zero doppler toward the doppler of the chaff returns as the relative level of the land clutter becomes small.

2.15 RADAR CLUTTER MAPS

In many MTI radar applications, the clutter-to-noise ratio in the receiver will exceed the improvement factor limit of the system even when techniques such as sensitivity time control (STC), improved radar resolution, and reduced antenna gain close to the horizon are used to reduce the level of clutter returns. The resulting clutter residues after the MTI canceler must, therefore, be further suppressed to prevent saturation of the PPI display and/or an excessive false-alarm rate in an automatic target detection (ATD) system.

Against spatially homogeneous sources of clutter such as rain, sea clutter, or corridor chaff, a cell-averaging constant-false-alarm-rate (CA-CFAR) processor following the MTI filter will usually provide good suppression of the clutter residues. Special features are sometimes added to the CA-CFAR, such as greatest-of-selection or two-parameter (scale and shape) normalization logic, in order to improve its effectiveness at clutter boundaries if the probability distribution of the clutter amplitude is non-gaussian. However, when the clutter returns are significantly nonhomogeneous, as is the case for typical land clutter returns, the performance of the cell-averaging CFAR will not be satisfactory and other means must be implemented to suppress the output residues to the noise level.

The traditional solution to this problem has been to deliberately reduce the receiver dynamic range prior to the MTI filter to the same value as the maximum system improvement factor. Theoretically, then, the output residue should be at or below the normal receiver noise level, and no false alarms would be generated. In practice, the introduction of IF limiting against the ground clutter returns will result in an additional

improvement factor restriction, as discussed in Section 2.11. Consequently, for the limited IF dynamic range to have the desired effect on the output residues, the limit level must be set 5 to 15 dB below the improvement factor limit of the linear system. The net result is that some of the clutter suppression capability of the MTI radar must be sacrificed in exchange for control of the output false-alarm rate.

Since returns from land clutter scatterers usually are spatially fixed and, therefore, appear at the same range and bearing from scan to scan, it has long been recognized that a suitable memory circuit could be used to store the clutter residues and remove them from the output residue on subsequent scans by either subtraction or gain normalization. This was the basic principle of the so-called area MTI, and many attempts have been made to implement an effective version of this circuit over an extended span of time. The main hindrance to its success has been the lack of appropriate memory technology, since the storage tube (long the only viable candidate) lacks in resolution, registration accuracy, simultaneous read-and-write capability, and stability. The development of high-capacity semiconductor memories is the technological breakthrough that has made the design of a working area MTI a reality. The *area MTI* is better known today as a *clutter map*, but both terms are used.

The clutter map may be considered as a type of CFAR where the reference samples, which are needed to estimate the level of the clutter (or clutter residue), are collected in the cell under test on a number of previous scans. Since aircraft targets usually move several resolution cells from one scan to the next, it is unlikely that the reference samples will be contaminated by a target return. Alternatively, by making the averaging time (in terms of past scans) long, the effect of an occasional target return can be minimized. Although the primary purpose of the clutter map is to prevent false alarms due to discrete clutter or clutter residues that are at a fixed location, it may also be necessary to consider slowly moving point clutter in the clutter map design, either to suppress bird returns or because the radar is on a moving platform (e.g., a ship).

The memory of a clutter map is usually organized in a uniform grid of range and azimuth cells, as illustrated in Figure 2.89. Each map cell will typically have 8 to 16 bits of memory so that it will handle the full dynamic range of signals at its input, which makes it possible to detect a strong target flying over a point of clutter (sometimes referred to as *superclutter visibility*). The dimensions of each cell are a compromise between the required memory and several performance characteristics. These are the minimum target velocity that will not be suppressed by the map (so-called cutoff velocity), its transient response, and the loss in sensitivity caused by the clutter map (similar to a CFAR loss). The minimum cell size will be constrained by the size of the radar resolution cell.

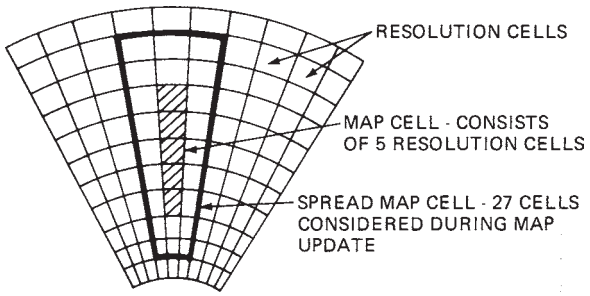


FIGURE 2.89 Clutter map cell definition

Each map cell is updated by the radar returns (or residues) falling within its borders (or in its vicinity) on several previous scans. To save memory, the cells are usually updated by using a simple recursive (single-pole) filter of the form

$$y(i) = (1-\alpha)\cdot y(i-1) + \alpha \cdot x(i) \tag{2.61}$$

where $y(i-1)$ is the clutter map amplitude from the previous scan, $y(i)$ is the updated clutter map amplitude, $x(i)$ is the radar output on the present scan, and the constant α determines the memory of the recursive filter. The test for detecting a target based on the output $x(i)$ is

$$x(i) \geq k_T \cdot y(i-1) \tag{2.62}$$

where the threshold constant k_T is selected to give the required false-alarm rate. Alternatively, the radar output can be normalized on the basis of the clutter map content to obtain an output $z(i) = \frac{x(i)}{y(i-1)}$, which can be processed further if required. Analogously to the implementation of the cell-averaging CFAR processor, the amplitude $x(i)$ can be obtained using a linear, square-law, or logarithmic detector.

The loss in detectability due to the clutter map is analogous to the CFAR loss analyzed in the literature for many different conditions. An analysis of the clutter map loss for single-hit detection using a square-law detector has been presented by Nitzberg.[44] These and other results can be summarized into a single universal curve of clutter map loss, L_{CM}, as a function of the clutter map ratio x/L_{eff}, as shown in Figure 2.90, where x defines the required false-alarm probability according to $P_f = 10^{-x}$ and L_{eff} is the effective number of past observations averaged in the clutter map defined as

$$L_{eff} = \frac{2-\alpha}{\alpha} \tag{2.63}$$

For example, for $P_f = 10^{-5}$ and $\alpha = 0.125$, the clutter map loss is $L_{CM} = 1.8$ dB since $x = 5$ and $L_{eff} = 15$ for this case. Also shown in Figure 2.90 is the curve for the conventional CA-CFAR,[45] where all reference samples are equally weighted. If more than one noise and/or clutter amplitude is used to update the clutter map content on each scan, the value of L_{eff} should be increased proportionally. It should also be noted that most radars base their target detection on multiple hits using some form of video integration, and that a clutter map loss based on the single-hit results of Figure 2.90 could be much too large.

An analysis of the performance of typical implementations of clutter maps has been discussed in Khoury and Hoyle.[46] From this reference, a typical transient-response curve is shown in Figure 2.91 for a single point clutter source 20 dB above thermal noise that fluctuates from scan-to-scan according to a Rayleigh probability density function, a filtering constant of $\alpha = 0.125$ and assuming four returns noncoherently integrated in each clutter map cell. The abscissa is in radar scans, and the ordinate is probability of detection of the point clutter source. Since the clutter point has the same amplitude statistics as thermal noise, the output false-alarm rate approaches $P_f = 10^{-6}$ asymptotically.

Against a slowly moving source of clutter (e.g., birds), the probability of detection may increase as the clutter source crosses the boundary between two clutter map cells. To prevent this, a spreading technique can be used, through which each clutter map cell will be updated—not only with radar returns falling within its boundaries, but also

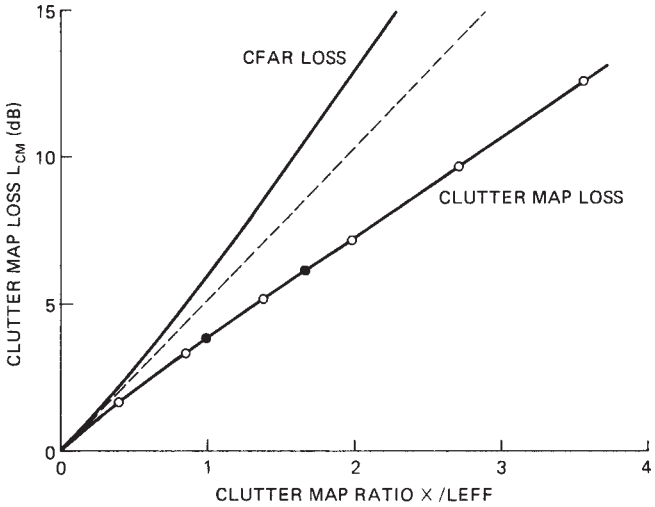


FIGURE 2.90 Universal curve for determining detectability loss caused by the clutter map

by using radar returns in adjacent cells in range and azimuth. Through the use of such spreading, an additional degree of control over the clutter map velocity response can be achieved.

An example of the velocity response of a clutter map including such spreading is shown in Figure 2.92. The range extent of the clutter map cell is 5 μs, the radar resolution cell is 1 μs, $n = 4$ pulses are noncoherently integrated, the filtering constant is $\alpha = 0.125$, the update interval is 5 s, and the $SNR = 20$ dB. On each scan, the clutter map cell is updated with the radar amplitudes in the five range cells falling within the clutter map cell and with the amplitude from one additional radar resolution cell before and after the clutter map cell.

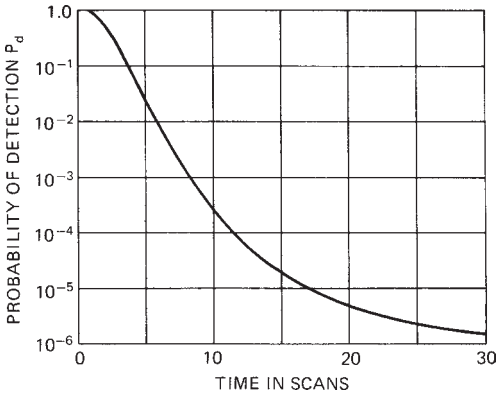


FIGURE 2.91 Transient response of clutter map due to Swerling Case 2 point clutter model

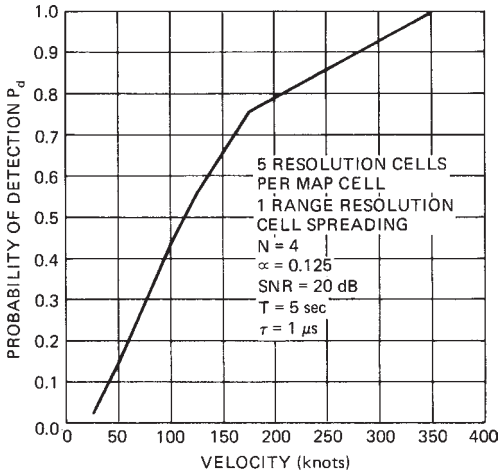


FIGURE 2.92 Velocity response of clutter map

It is seen from Figure 2.92 that the velocity response characteristic of the clutter map from stopband to passband is somewhat gradual in this particular implementation. This is partly due to the large size of the clutter map cell relative to the radar resolution. A finer-grain map with additional spreading would have a much better velocity response characteristic.

A potential problem with the type of amplitude clutter map described in this section is the fact that a large target flying in front of a smaller target may cause enough buildup in the map to suppress the small target. One way to overcome this problem in a system that includes automatic tracking would be to use the track prediction gate to inhibit updating of the clutter map with new (target) amplitudes.

2.16 SENSITIVITY-VELOCITY CONTROL (SVC)

In the mid-1980s, several radar researchers had realized that signal processing algorithms to estimate the unambiguous radial velocity of a target using multiple PRF dwells during the time of target were becoming practical. These radial velocity estimates could be used for improved false-alarm control against slow-moving targets such as birds.[30,47] When such radial velocity measurements are paired with corresponding cross section estimates a powerful discriminant for distinguishing between slow-moving birds and low cross-section missiles becomes possible using the so-called sensitivity velocity control (SVC) algorithm.[48]

The SVC Concept. Sensitivity velocity control (SVC) is used when a radar must detect aircraft and missile targets in the presence of returns from unwanted targets such as large birds or bird flocks. The criteria to accept or reject targets is based on a combination of the radial velocity and apparent RCS (radar cross section) of the target returns. The desired targets may have an RCS smaller than a single bird, or possibly

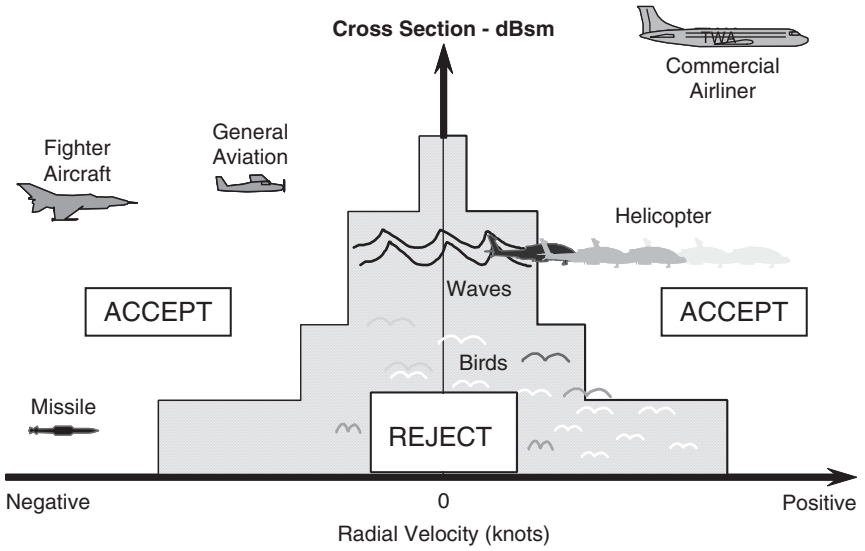


FIGURE 2.93 Illustrative acceptance/rejection criteria of SVC

a bird flock (in a single radar resolution cell). Thus, discrimination requires a parameter in addition to the target RCS. The available parameter is target radial velocity. Birds typically fly at 40 knots or less, whereas targets of concern usually have airspeeds of 100 knots or more. If the radar can make unambiguous radar doppler measurements of, e.g., ±160 knots with a single CPI (coherent processing interval), the radar can determine the true radial velocity of each radar echo from returns of three or more consecutive CPIs at different PRFs.

The acceptance criteria of the SVC algorithm[48] relates to the type of target (aircraft, missile, bird, etc.) being accepted or rejected. In general, the criteria accepts large targets having low to high radial velocities. The smaller the apparent radar cross section of the target, the higher the true radial velocity must be for acceptance. The true radial velocity versus apparent radar cross section profile is intended to accept aircraft and missiles but reject birds. Therefore, threatening targets that have high radial velocities, but very small RCS, can be instantly identified, whereas returns from birds, with their slow radial velocities, can be censored. A typical SCV accept/reject algorithm is depicted in Figure 2.93.

To obtain the doppler space of ±160 knots, ambiguous range PRFs must be used. This requires approximate PRFs of 1400 Hz at L band, 3,300 Hz at S band, and 11,000 Hz at X band (unambiguous ranges, respectively, 58 nmi, 27 nmi, and 5 nmi). The tradeoff for selecting PRFs is that in a dense target environment, when trying to resolve true radial velocity using different PRFs, "ghosts"[§] may be created.

[§] "Ghosts" occur when targets (or noise peaks) at different unambiguous ranges fold into the same, but incorrect, true range cell. The velocity resolution algorithm then gives an incorrect result, and the ghosts may be declared as threatening targets.

In addition to the "ghost" problem, multiple range ambiguities lead to targets having to compete with clutter at all ranges. In particular, targets at long distances have to compete with strong clutter returns in the first, or several, range intervals.

Because of the ghosting problem, in order to minimize range ambiguities while retaining adequate doppler space, RF frequencies of 1400 MHz or lower are best suited for the SVC unwanted target discrimination technique.

Range- and Range Rate Ambiguity Resolution. To apply the SVC algorithm, true range and radial velocity (range-rate) must be determined from the range-ambiguous and doppler-ambiguous waveform. This requires multiple detections from the same target. Assume a doppler filter bank of n-pulse FIR filters and assume a processing dwell that consists of three CPIs. The CPIs must use different PRFs and may also employ different RF frequencies. (The different RF frequencies change target RCS statistics from Swerling 1 to Swerling 2, and thus less radar energy is required for high probability of detection.) The CPIs must have (1) sufficient transmitted pulses so that n returns (enough to fill an n-pulse filter) will be received from the most distant target of interest and the most distant clutter and (2) one additional pulse to enable velocity determination (more on this later).

True Range Determination. The most straightforward way to detect a target and simultaneously determine its true range is to determine, on each CPI, all "primitive" detections at the output of the doppler filter bank. For this, it is assumed that each doppler filter output is processed through an appropriate clutter map threshold and cell-averaging CFAR to control the false-alarm rate. For each peak detection, adjacent amplitudes will be used to obtain an accurate ambiguous range estimate denoted $\hat{r}_i$, where the subscript refers to the CPI number. Also, from the specific doppler filter corresponding to the peak detection described above, the phase (θ_{1i}) of the return is saved. In addition, a corresponding phase (θ_{2i}) obtained from an identical second doppler filter bank trailing (or leading) the detection filter bank by one pulse repetition interval (PRI) is saved. This explains why a CPI of $n + 1$ pulses is needed to implement the SVC concept. For each primitive detection in a CPI, calculate the set of all possible target ranges out to the maximum instrumented range $R_{\max}$:

$$\begin{aligned} &\hat{R}_i = \hat{r}_i + m \cdot R_{\text{PRI},i} \qquad m = 0,1,2,...,m_{\max} \\ &\text{where } m_{\max} = \text{int}(R_{\max} / R_{\text{PRI},i}) + 1 \qquad i = 1,2,3 \end{aligned} \tag{2.64}$$

where $R_{\text{PRI},i}$ is the ambiguous range interval corresponding to the ith CPI. After the primitive detections from all CPIs in the processing dwell have been processed, the values of $\hat{R}_i$ from all CPIs are sorted into a single list. A final range detection and its true range is then found as a cluster of three primitive detections having possible ranges within an error window of two to three times the standard deviation of the ambiguous range estimate.

True Radial Velocity Determination. For each true target detection, an unambiguous radial velocity estimate must next be determined using a similar procedure to that described above for range. For this, an accurate estimate, $f_{d,i}$, of the ambiguous target radial velocity must be obtained at the range corresponding to the ambiguous primitive target detection on each CPI. This frequency estimation problem has been studied by many authors with the best approach being defined

by the maximum likelihood estimate.[49] For a single-pulse signal-to-noise ratio S_1 and n pulses in a CPI, the Cramer-Rao lower bound for the accuracy of the doppler frequency estimate is

$$\frac{\sigma_f}{\text{PRF}} = \frac{\sqrt{6}}{2 \cdot \pi \cdot \sqrt{S_1 \cdot n \cdot (n^2 - 1)}} = \frac{0.3898}{\sqrt{S_1 \cdot n \cdot (n^2 - 1)}} \tag{2.65}$$

Since the maximum likelihood estimation procedure tends to require a tedious computational burden, a simplified approach for estimating the doppler frequency is highly desirable. One such approach using phase measurements of the doppler filter output at times separated by one interpulse period* was presented in McMahon and Barrett.[50] The normalized doppler frequency estimate is

$$\frac{f_{d,i}}{\text{PRF}} = \frac{\theta_{1,i} - \theta_{2,i}}{2 \cdot \pi} \tag{2.66}$$

and the corresponding radial velocity is

$$\hat{v}_i = \frac{f_{d,i} \cdot \lambda}{2} \tag{2.67}$$

In most cases of interest, the accuracy of this estimate of doppler frequency is as good as the maximum likelihood procedure. Expressed in terms of the numerator of Eq. 2.65, which will be denoted by k, a simulation of the phase-difference estimator using different weighting functions for the doppler filter bank are summarized in Figure 2.94. It is noted that the performance of the phase-difference estimation procedure is best when moderate Taylor weighting functions are used. For uniform weighting, the procedure would be substantially inferior to the maximum likelihood approach. The increase in the constant k for the more severe weighting cases is the result of the SNR loss resulting from the use of weighting.

Using an approach similar to that used to resolve the range ambiguity, all possible radial velocities are then enumerated to the maximum negative and positive radial velocity of interest on each of the CPIs:

$$\begin{aligned} &\hat{V}_i = \hat{v}_i + m \cdot V_{B,i} \qquad m = -m_{\max}, -(m_{\max} - 1), ..., 0, 1, 2, ..., m_{\max} \\ &\text{where } m_{\max} = \text{int}(V_{\max} / V_{B,i}) + 1 \qquad i = 1, 2, 3 \end{aligned} \tag{2.68}$$

In this equation, $V_{B,i} = \text{PRF}_i \cdot \lambda / 2$ is the blind velocity for the ith CPI. The possible target radial velocities for all CPIs are then sorted into a single list, and the most likely true radial velocity is found where at least two possible velocities fall within an interval less than two or three times the standard deviation of the doppler frequency estimate. The tightness of the cluster of nearly identical velocities in conjunction with the number of CPIs contributing to the cluster can be utilized as a measure of reliability of the unambiguous radial velocity estimate.

* This approach was first brought to the attention of the authors by Dr. Ben Cantrell of the U.S. Naval Research Laboratory.

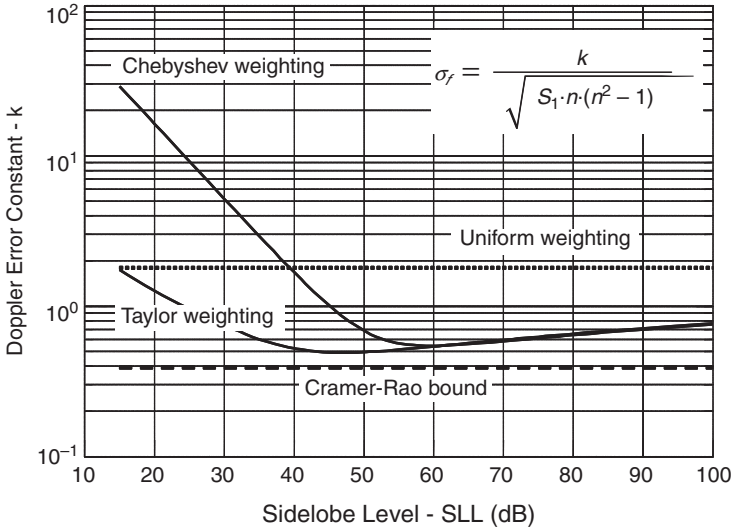


FIGURE 2.94 Performance of phase-difference doppler frequency estimator for different weighting functions of the doppler filter bank

Comments. The above procedure for determining true range and true radial velocity has been described for a dwell of three CPIs and the assumption that each target will have a return for each of the three CPIs. In practice, this assumption is not always valid, and the actual implementation may choose, for example, to have the dwell consist of four or five CPIs, with the range and velocity determinations being based on the best grouping of three returns. The actual implementation must be based on the parameters of the system and permissible time allocated for each dwell.

The PRFs of the CPIs should be selected to minimize the chance of false radial velocity determinations. One method of selecting PRFs is similar to selecting pulse interval ratios for staggered PRF operation, as described in Section 2.8. For example, if operating at an average RF frequency of 1300 MHz, at an average PRF of 1400 Hz (ambiguous velocity of 312 knots), and covering a velocity range of interest of ±2500 knots, there are approximately 16 doppler ambiguities to cover. Using the factors of –3, 2, –1, 3, as used in PRF stagger selection, the interpulse periods of the four different PRFs would be in the ratio of 13, 18, 15, 19. The average of these ratios is 16.25. The PRFs are calculated as 16.25·1400/13, 16.25·1400/18, 16.25·1400/15, and 16.25·1400/19. The PRFs would be about 1750, 1264, 1517, and 1197 Hz.

2.17 CONSIDERATIONS APPLICABLE TO MTI RADAR SYSTEMS

MTI radar system design encompasses much more than signal processor design. The entire radar system—transmitter, antenna, and operational parameters—must be keyed to function as part of an MTI radar. For example, excellent MTI concepts will not perform satisfactorily unless the radar local oscillator is extremely stable and the

transmitter has very little pulse-to-pulse frequency or phase jitter. In addition, the system must successfully operate in an environment that comprises many unwanted targets, such as birds, insects, and automobiles.

Hardware Considerations. In this section, rules and facts relating to MTI radar design, as developed during many years of work in the field, will be summarized:

The rules are as follows:

1. Operate at constant duty cycle.
2. Synchronize ac-dc and dc-dc power conditioners† to harmonics of the PRF.
3. Design the system to be fully coherent.‡
4. Provide IF Limiters prior to A/D converters.
5. Be wary of vibration and acoustic noise.

The facts are as follows:

1. The basic MTI concept does not require a long time on target to resolve targets from fixed clutter. Instead, MTI systems reject fixed clutter through a subtraction process while retaining moving targets.
2. Transmitter intrapulse anomalies have no affect on MTI performance if they repeat precisely pulse-to-pulse.

Rule 1. Operate at constant duty cycle. The transmitter (whether the transmitter is a single large tube or a distributed function as in an active phased array with many transmit-receive elements) should be operated at constant duty cycle. This permits the transmitter power supply transient effects to be identical pulse to pulse and also, particularly applicable to solid-state transmit devices, permits the device heating and cooling to be identical from pulse to pulse. Sometimes constant duty cycle operation is not possible, but there are various techniques that can be used to approach this desired condition. Consider an MTD waveform where a CPI consisting of n pulses is transmitted with a constant PRI. The next CPI uses a different PRI. Constant duty cycle can be maintained by changing the transmitted pulse length in proportion to the change in the PRI. If pulse compression is used, the range resolution of the compressed pulse can be maintained by changing the pulse compression waveform. If it is necessary to utilize precisely the same waveform and RF pulse length from CPI to CPI, with, for example, a klystron transmitter, the beam pulse of the klystron can be varied to maintain constant beam duty cycle while the RF pulse length is maintained constant. This wastes part of the beam pulse energy for the longer PRIs, but the average power loading on the power supply remains constant. The same technique can be utilized with solid-state devices by changing the drain voltage pulse duration, while holding the RF pulse constant. A second-order correction that has been utilized when changing between CPIs with different PRIs is to have a transition PRI that is the average of the two PRIs. With phased array radars, if the beam transition time between CPIs takes longer than a PRI, it is important to keep the transmitter pulsing at a constant duty cycle during the transition time. If constant duty cycle cannot be maintained, or when starting to radiate

† Power conditioners accept either ac or dc input and provide a regulated dc output.

‡ "Fully coherent" is described under rule 3.

after dead time, the transmitter, power supply, and heating effects must be allowed to settle before good MTI performance can be expected. The duration of the settling time depends on the system parameters and the requirements.

Rule 2. Synchronize ac-dc and dc-dc power conditioners to harmonics of the PRF. When ac-dc and/or dc-dc power conditioners are used for voltages applied to transmitting devices, the frequency (and its harmonics) of the converter must be attenuated sufficiently so that they do not modulate the phase of the transmitted pulses. If the power conditioner frequencies cannot be sufficiently attenuated, their frequency should be synchronized to a multiple of the PRF of the CPI so that modulations repeat precisely pulse-to-pulse and thus will cancel like stationary clutter.

Rule 3. Design the system to be fully coherent. All frequencies and timing signals should be generated from a single master oscillator. Doing this makes the entire system coherent, and mixer products will be identical pulse-to-pulse and will, therefore, cancel in the MTI filters. When this coherence of all frequencies is not maintained, clutter residue will occur and must be quantified to determine if it is at an acceptable level. One of the prominent places in which residue caused by unsynchronized local oscillators has shown up is in pulse-compression sidelobes. If the pulse-compression sidelobes from fixed clutter returns vary from pulse to pulse, they do not cancel. This coherency issue has been further discussed by Taylor.[51]

Rule 4. Provide IF Limiters prior to A/D converters. MTI radars require that IF bandpass limiters exist prior to an A/D (analog/digital converter). The limiter prevents any clutter return from exceeding the dynamic range of the A/D. This requirement exists for either quadrature I, Q (in-phase, quadrature) sampling or direct sampling with the I and Q data constructed after the A/D. The limiter must be designed to minimize the conversion of amplitude to phase no matter how much the signal level exceeds the limit level. If clutter saturates the A/D, the I, Q data is significantly corrupted. When limiters prevent A/D saturation, the signals are limited in a controlled manner that still enables good clutter rejection about 90% of the time.

Rule 5. Be wary of vibration and acoustic noise. Many RF devices are susceptible to both vibration and acoustic noise. An air conditioner fan blowing on waveguide has caused degradation of improvement factor due to phase modulation of signals. Vibrations can cause phase modulation of an oscillator. Acoustic noise can originate from cooling fans, and vibrations can come from shipboard or airborne radar platforms. Components such as klystrons and solid-state modules can have unexpected susceptibility to vibration. RF connectors must be secure. Shock mounts can be used to isolate components from the cabinet structure. It is recommended that all RF components, in their operational configuration, be tested for phase stability in the vibration environment in which they will be used.

Fact 1. The basic MTI concept does not require sufficient time-on-target to resolve targets from fixed clutter using a linear time-invariant filter. Instead, MTI systems reject fixed clutter through a subtraction process while retaining moving targets. An MTI system using a two-pulse canceler requires the transmitter to transmit only two successive, identical pulses for the system to be able to reject stable fixed clutter. The radar returns from the second pulse are subtracted from the returns from the first pulse.

The result from this subtraction process is that the fixed clutter is removed, and moving targets are retained. The output from the first pulse is not used, making this type of MTI filter time-variant. Of course, the clutter filters may be more complex than a two-pulse canceler,[§] but the principle still remains that fixed clutter is rejected by the zeros in the canceler transfer characteristic. This enables phased array radars to have good clutter rejection with short dwells.[*]

Fact 2. Transmitter intrapulse anomalies have no affect on MTI performance if they repeat precisely pulse to pulse. Transmitted pulses should be identical. It does not matter if there is intrapulse amplitude or frequency modulation of the transmitted pulse, as long as it repeats precisely from pulse to pulse. If the voltage of the transmitter power supply varies pulse to pulse, the transmitted pulses will not be identical, and the resulting variations must be quantified to determine if the limitations on improvement factor fall within the stability budget for the system. However, if the only difference between pulses is absolute phase (not intrapulse variations pulse to pulse), some mitigation is possible. One method of compensating for small variations in the phase of transmitter pulses follows. Lincoln Laboratory changed the original TDWR waveform to an MTD type waveform. (The original TDWR waveform was constant PRF during each antenna rotation, and processing was done with elliptic filters.) They then modified the system "…to achieve 65-dB clutter suppression using a nearby water tower for a target."[52] The TDWR uses a klystron transmitter tube. Typical phase pushing for a klystron due to modulator voltage change is 10° for 1% delta-E/E. The stability budget allocated a 75-dB limit on improvement factor to the transmitter, and this required that the rms pulse-to-pulse power supply voltage variation be less than 1 part in 100,000. The transmitter power supply could not meet this requirement when the radar changed PRF from CPI to CPI, as required by an MTD waveform. Therefore, the actual phase of each transmitted pulse was measured, and this measured value was used to correct the phase of the received signals for that PRI. This technique causes small perturbations in phase from weather signals received from ambiguous ranges, but does not interfere with velocity estimates. (It does degrade the improvement factor of clutter signals received from ambiguous ranges, but for the TDWR operation, that degradation was deemed acceptable.)

Environmental Considerations. This discussion contains essential information for those designing a modern surveillance radar to detect man-made airborne targets. The laws of physics combined with the environment make it impossible to design an MTI surveillance radar that does not have compromises. The problems are related to the unwanted returns from birds, insects, automobiles, long-range fixed clutter, and short- and long-range weather.[53] The current state-of-the-art of radar can ameliorate these problems, but not without some undesirable side effects. (Many unwanted point target returns have characteristics similar to the returns from wanted targets, and the unwanted returns may outnumber returns from desired targets by the thousands.)

[§] The clutter filters must be designed based on system parameters to reject the radial speed of the "fixed" clutter. See Sections 2.4 and 2.6.

[*] It has been observed that some phased array radars have poor clutter rejection, which is often caused by failure to follow rule 1.

The problems are exacerbated when anomalous or ducted propagation occurs (*anomalous propagation,* as used herein, is when the radar energy follows the curvature of the Earth, thus causing detection of both fixed and moving clutter at long ranges). Figure 2.95 from Shrader[53] shows PPI photographs taken with an ARSR-2 radar mounted on a 50-ft tower in flat country near Atlantic City, New Jersey. With normal propagation, the expected line-of-sight is about 10 nmi, but the clutter actually goes out to 100 nmi. The bridges across the intracoastal waterway can be seen. On occasion, the unwanted long-range clutter and weather returns come from ambiguous ranges.

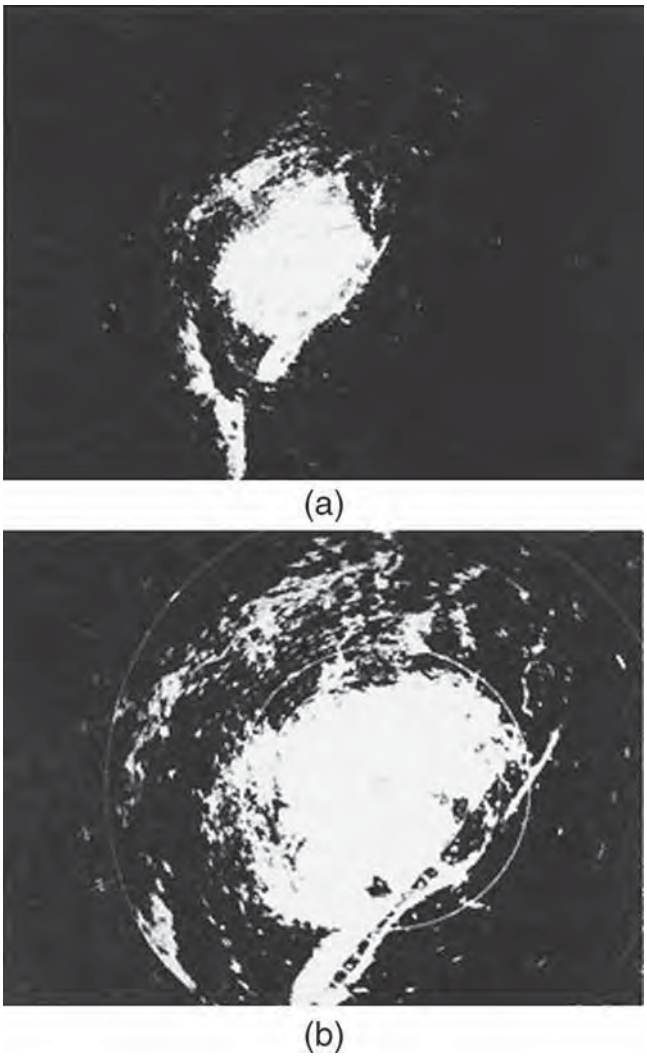


FIGURE 2.95 Anomalous propagation (ducting): (*a*) 100-nmi maximum range and (*b*) 50-nmi maximum range

The radar system must have features to cope with these situations. For example, if pulse-to-pulse staggering is used, the ambiguous-range clutter will not cancel and either the PRI must be increased or the PRI must be made constant over the azimuth angles from which the ambiguous range clutter is received. And be forewarned of a pitfall into which many radar designers have fallen. For example, when presented with the requirement to track 20 targets, the designer may not realize that radar returns from the 20 targets of interest may be embedded in similar returns from thousands of unwanted targets.

A typical long-range air-traffic-control radar has sufficient sensitivity to detect a single large bird, such as a crow, seagull, or vulture (approximate RCS of 0.01 square meter) at a range of 50 miles. If there are many such birds in the resolution cell of the radar, then the composite RCS increases. Ten large birds in a resolution cell will have an RCS of 0.1 square meter. When multipath reflections occur, such as over the ocean when the radar beam is centered at the horizon, there can be up to a 12 dB enhancement of the RCS of the birds, giving an apparent RCS greater than one square meter to the flock of 10 birds. If there is 1 bird (or bird flock) per square mile, there will be about 3000 bird returns within 30 miles of the radar.

Techniques used to counter unwanted targets are as follows:

1. Sensitivity time control (STC) used for eliminating low RCS targets in low PRF radars—that is, radars that have no range ambiguities under normal operation
2. Enhanced high-angle gain antennas
3. Two-beam antennas—beam lifted above the horizon for short-range reception, and then lowered to horizon for long range
4. MTD techniques using clutter maps. Also counting detections in small range-azimuth sectors and increasing detection thresholds in each sector if too many detections occur.
5. PRFs high enough so that all targets with radial velocities below 40 knots can be censored
6. Sensitivity velocity control (SVC), which censors radially slow, small targets, while accepting radially fast targets and large targets

Combinations of techniques 1 through 4 are used in most air-traffic-control radars where the smallest targets of interest have an RCS of one-square meter or greater. Techniques 5 and 6 are used when the desired targets may have radar cross sections similar to, or smaller than, a bird.

Technique 1. STC is the traditional method of suppressing birds and insects in a radar with an unambiguous range PRF (a PRF low enough so that the range to targets and clutter is unambiguous). STC decreases the sensitivity of the radar at short range and then increases sensitivity, usually using a fourth-power law, as range increases. This has the effect of not permitting detection of targets with apparent radar cross sections of, say, less than 0.1 square meter. Figure 2.96 shows how effective STC can be against birds. These PPI photos were taken with an L band ARSR (air-route surveillance radar) in Oklahoma. Note that the majority of returns from birds were eliminated, but not 100%. Figure 2.97 shows the effect of STC against bats and insects.[†]

[†] Daytime bird returns and nighttime bat and insect returns can often be seen in real time—the extent depends on the weather and time of year—on the NEXRAD (WSR-88D) weather radar images on the NOAA Internet sites.

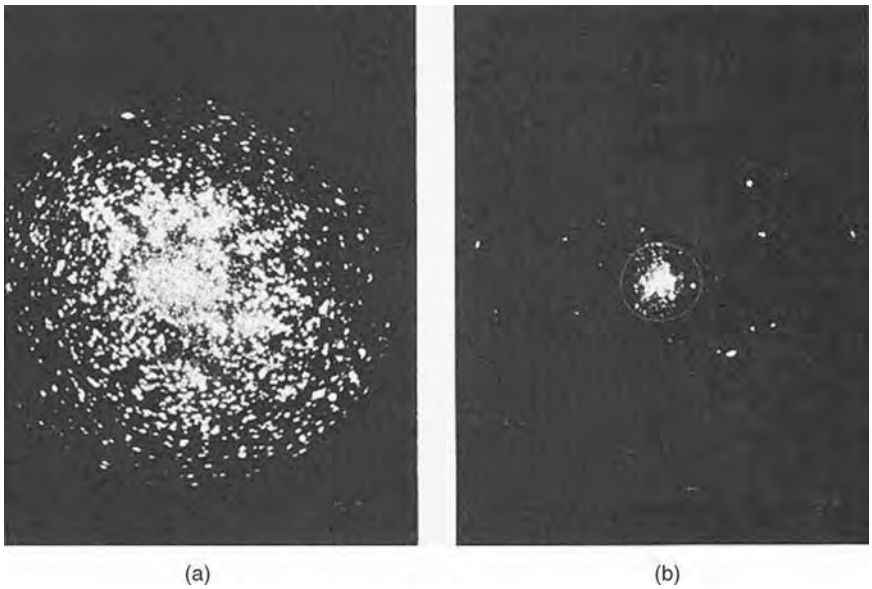


FIGURE 2.96 STC can greatly reduce the number of birds displayed. Range 25 nmi. (*a*) Birds seen with MTI and (*b*) birds seen with MTI and STC.

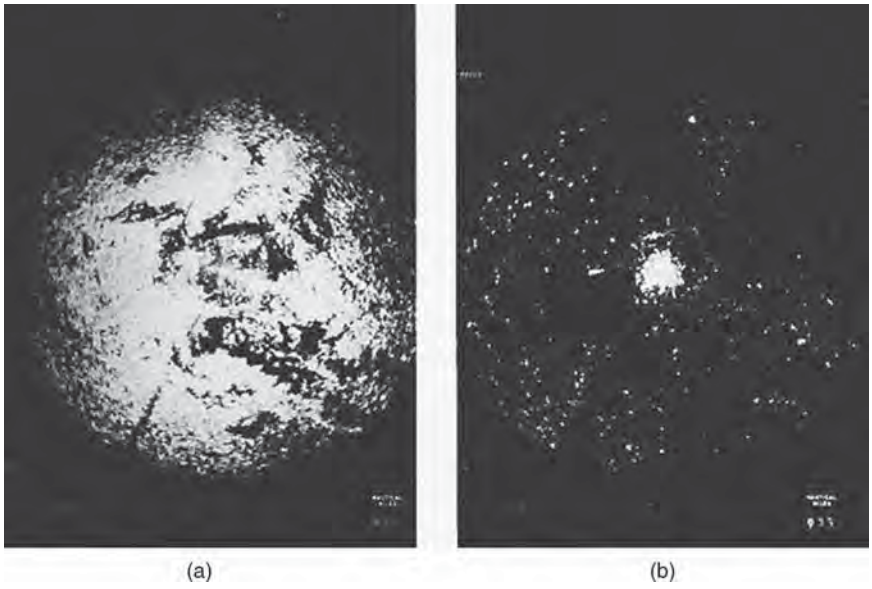


FIGURE 2. 97 Insects with and without STC and range 25 miles: (*a*) bats and insects seen with MTI and (*b*) bats and insects seen with MTI and STC

The typical doppler radar images presented by TV weather forecasters often have the birds and bats and insects removed by human intervention.

Technique 2. STC works quite well for unwanted biological returns near the peak of the radar beam, but when used with a cosecant-squared antenna beam it solves one problem but creates another: it also decreases sensitivity to desired targets at high elevation angles where the antenna gain is low. The solution to this problem is to boost the antenna gain at high elevation angles to be considerably higher than the requirement for the cosecant-squared pattern. Not only does this compensate for the use of STC, but also enhances the target-to-clutter signal ratio for targets at high elevation angles, thus improving MTI performance. The penalty for this solution is a loss in the peak antenna gain that can be achieved. An illustration of this approach is provided in Figure 2.98, which shows both the ARSR-2 antenna pattern and the corresponding free-space coverage.

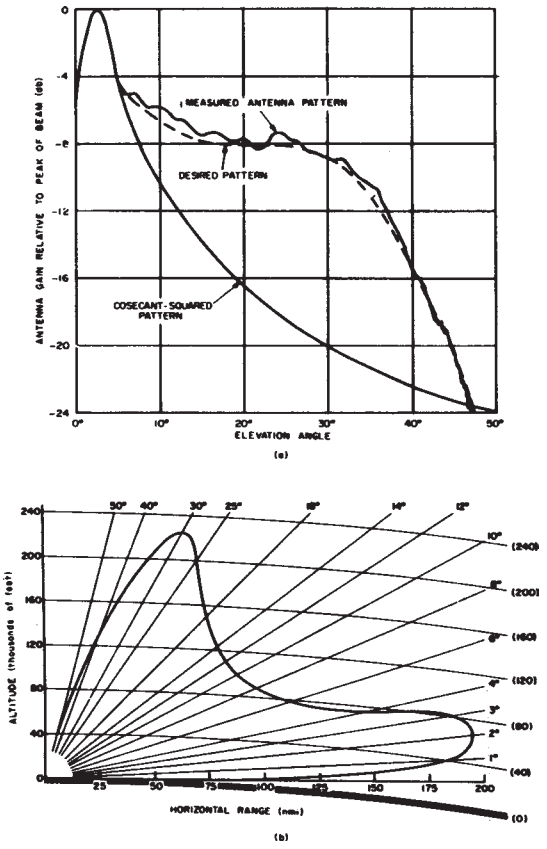


FIGURE 2.98 Antenna elevation pattern for the ARSR-2 antenna: (*a*) compared with the cosecant-squared pattern and (*b*) free-space coverage diagram

The loss in peak gain for this example, due to the boost of coverage at high angles, was about 2 dB. The combination of STC with enhanced high-angle coverage does quite well for insects and birds, but does not eliminate automobile and truck returns. Vehicles have RCSs that equal or exceed the RCS of many desired aircraft targets.

Technique 3. The two-beam technique reduces the returns from very low elevation angles where vehicle traffic (and many birds, bats, and insects) is encountered. The radar transmits energy using the basic pattern, but uses a higher angle beam for reception at shorter ranges, and the basic antenna pattern for receiving at longer ranges. Figure 2.99 shows, underneath the transmitting feed horn, a second receive-only antenna feed horn for the high beam. The effective two-way antenna patterns are shown in Figure 2.100.

As previously mentioned, the above techniques (STC, two-beam antennas, and some variation of MTD) are currently used on many air-traffic-control radars. The two-beam antennas also utilize some high-angle gain enhancement to counter the high-angle effects of STC.

Technique 4. The MTD approach is described in Section 2.2.

Technique 5. A brute-force technique used to eliminate targets with radial velocities of less than approximately ±40 knots resulting in a total rejection interval of 80 knots. To keep this rejection of velocities to no more than 25% of the doppler space available, the ambiguous velocity must be about 320 knots. This requires PRFs of 1,400 Hz at L band, 3,300 Hz at S band, and 11,000 at X band (unambiguous ranges, respectively, 58 nmi, 27 nmi, and 5 nmi). The main challenge with this technique is that fixed clutter returns from many range ambiguities, as well as all targets of interest, fold into the first range interval. Thus, excellent clutter rejection must be provided to prevent folded clutter from suppressing targets of interest, which may be at any true range.

Technique 6. SVC, as described in Section 2.16, is used when it is necessary to distinguish very low RCS targets from low velocity clutter, such as birds, insects, and sea. Somewhat lower PRFs can be used than those used for technique 5 because the

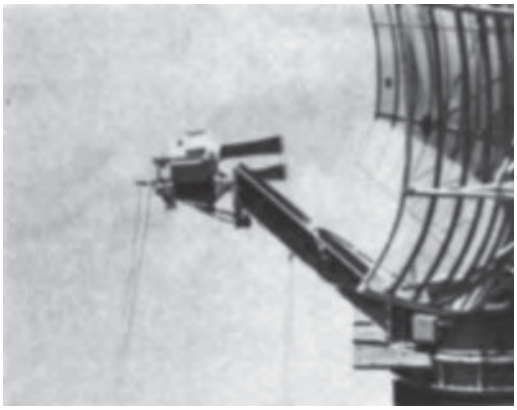


FIGURE 2.99 Two-beam antenna

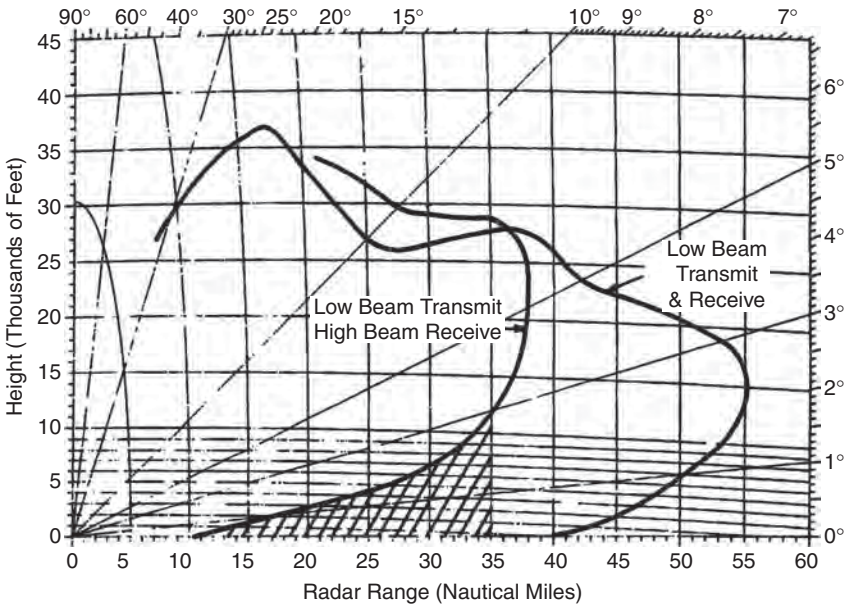

90° 60° 40° 30° 25° 20° 15° 10° 9° 8° 7°
Height (Thousands of Feet)
Radar Range (Nautical Miles)
Low Beam Transmit & Receive
Low Beam Transmit High Beam Receive
0 5 10 15 20 25 30 35 40 45 50 55 60
0 5 10 15 20 25 30 35 40 45
0° 1° 2° 3° 4° 5° 6°

FIGURE 2.100 Example of coverage obtained with a two-beam antenna

logic permits retaining many of the targets with smaller radial velocities if their RCS is large enough. SVC still rejects bird clutter, but retains, for example, the fast incoming, threatening low-RCS missile, while also retaining the larger cross-section aircraft with lower radial velocities.

REFERENCES

1. S. Applebaum, "Mathematical description of VICI," General Electric Co., Syracuse, NY, Report No. AWCS-EEM-1, April 1961.
2. S. M. Chow, "Range and doppler resolution of a frequency-scanned filter," *Proc. IEE*, vol. 114, no. 3, pp. 321–326, March 1967.
3. C. E. Muehe, "New techniques applied to air-traffic control radars," *Proc. IEEE*, vol. 62, pp. 716–723, June 1974.
4. R. J. Purdy et al., "Radar signal processing," *Lincoln Laboratory Journal*, vol. 12, No. 2, 2000.
5. R. J. McAulay, "A theory for optimum MTI digital signal processing," MIT Lincoln Laboratory, Lexington, MA, Report no.1972-14 Part I, Part II, Supplement I, February 22, 1972.
6. E. J. Barlow, "Doppler radar," *Proc. IRE,* vol. 37, pp. 340–355, April 1949.
7. W. L. Simkins, V. C. Vannicola, and J. P. Royan, "Seek Igloo radar clutter study," Rome Air Development Center, Report No. Rept. TR-77-338 (DDC AD-A047 897), October 1977.
8. W. Fishbein, S. W. Graveline, and O. E. Rittenbach, "Clutter attenuation analysis," US Army Electronics Command, Fort Monmouth, NJ, Report No.ECOM-2808, March 1967.
9. J. B. Billingsley, *Low-Angle Radar Land Clutter—Measurements and Empirical Models*, Norwich, NY: William Andrew Publishing, 2002.

10. F. E. Nathanson and J. P. Reilly, "Radar precipitation echoes—Experiments on temporal, spatial, and frequency correlation," The Johns Hopkins University, Applied Technology Laboratory, Report No. Tech. Memo TG-899, April 1967.
11. D. K. Barton, *Radar System Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1964.
12. D. K. Barton, "Radar equations for jamming and clutter," in *Supplement to IEEE Trans.* AES-3, EASCON'67 Tech. Conv. Rev, November, 1967, pp. 340–355.
13. R. J. Doviak and D. S. Zrnic, *Doppler Radar and Weather Observations*, Orlando, FL: Academic Press, 1984.
14. H. R. Ward, "A model environment for search radar evaluation," in *EASCON '71 Convention Record*, New York, 1971, pp. 81–88.
15. IEEE, "IEEE Standard Radar Definitions," Radar Systems Panel, IEEE Aerospace and Electronics Systems Society, Report No. IEEE Std 686-1997, 1999.
16. D. K. Barton and W. W. Shrader, "Interclutter visibility in MTI systems," in *IEEE EASCON '69 Tech. Conv. Rec*., New York, NY, October 1969, pp. 294–297.
17. D. K. Barton, *Modern Radar System Analysis*, Norwood, MA: Artech House, 2005, pp. 228–230.
18. L. Spafford, "Optimum radar signal processing in clutter," *IEEE Trans.*, vol. IT-14, pp. 734–743, September 1968.
19. L. A. Wainstein and Y. D. Zubakov, *Extraction of Signals From Noise*, New York: Dover, 1970.
20. J. Capon, "Optimum weighting functions for the detection of sampled signals in noise," *IRE Trans. Information Theory,* vol. IT-10, pp. 152–159, April 1964.
21. L. R. Rabiner et al., "Terminology in digital signal processing," *IEEE Trans. on Audio and Electroacoustics*, vol. AU-20, no. 5, pp. 322–337, December 1972.
22. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2001, p. 117.
23. H. Urkowitz, "Analysis and synthesis of delay line periodic filters," *IRE Trans. Circuit Theory*, vol. CT-4, no. 2, pp. 41–53, June 1957.
24. W. M. Hall and H. R. Ward, "Signal-to-noise ratio loss in moving target indicator," *Proc. IEEE*, vol. 56, pp. 233–234, February 1968.
25. W. W. Shrader and V. Gregers-Hansen, "Comments on 'Coefficients for feed-forward MTI radar filters'," *Proc IEEE*, vol. 59, pp. 101–102, January 1971.
26. W. D. White and A. E. Ruvin, "Recent advances in the synthesis of comb filters," in *IRE Nat. Conv. Rec.* vol. 5, pt. 2, New York, NY, 1957, pp. 186–200.
27. R. H. Fletcher and D. W. Burlage, "Improved MTI performance for phased array in severe clutter environments," in *IEEE Conf. Publ. 105*, 1973, pp. 280–285.
28 A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975, p. 223.
29. L. Zverev, "Digital MTI radar filters," *IEEE Trans.,* vol. AU-16, pp. 422–432, September 1968.
30. Ludloff and M. Minker, "Reliability of velocity measurement by MTD radar," *IEEE Trans.*, vol. AES-21, pp. 522–528, July 1985.
31. W. W. Shrader, "MTI Radar," Chap. 17 in *Radar Handbook*, M. I. Skolnik (ed.), New York: McGraw-Hill, 1970, pp. 17–19.
32. T. M. Hall and W. W. Shrader, "Statistics of clutter residue in MTI radars with IF limiting," in *IEEE Radar Conference*, Boston, MA, April 2007, pp. 01–06.
33. G. Grasso, "Improvement factor of a nonlinear MTI in point clutter," *IEEE Trans.*, vol. AES-4, November 1968.
34. H. R. Ward and W. W. Shrader, "MTI performance degradation caused by limiting," in *EASCON '68 Tech. Conv. Rec.,* (supplement to *IEEE Trans.* vol. AES-4), November 1968, pp. 168–174.
35. G. Grasso and P. F. Guarguaglini, "Clutter residues of a coherent MTI radar receiver," *IEEE Trans.*, vol. AES-5, pp. 195–204, March 1969.
36. T. A. Weil, "Applying the Amplitron and Stabilotron to MTI radar systems," in *IRE Nat. Conv. Rec.*, vol. 6, pt. 5, New York, NY, 1958, pp. 120–130.

37. T. A. Weil, "An introduction to MTI system design," *Electronic Progress*, vol. 4, pp. 10–16, May 1960.
38. D. B. Leeson and G. F. Johnson, "Short-term stability for a doppler radar: Requirements, measurements, and techniques," *Proc. IEEE*, vol. 54, pp. 329–330, February 1966.
39. Hewlett Packard Product Note 11729B-1, March 1984.
40. R. Vigneri et al., "A graphical method for the determination of equivalent noise bandwidth," *Microwave Journal*, vol. 11, pp. 49–52, June 1968.
41. D. Steinberg, "Chapters," Chaps.1–4 in *Modern Radar: Analysis, Evaluation, and System Design*, R. S. Berkowitz (ed.), New York, NY: John Wiley and Sons, 1966.
42. J. R. Klauder, "The theory and design of chirp radars," *Bell System Technical Journal*, vol. XXXIX, no. 4, pp. 745–809, July 1960.
43. W. Rice and K. H. Wu, "Quadrature sampling with high dynamic range," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-18, no. 4, pp. 736–739, November 1982.
44. R. Nitzberg, "Clutter map CFAR analysis," *IEEE Trans.*, vol. AES-22, pp. 419–422, July 1986.
45. V. Gregers Hansen, "Constant false alarm rate processing in search radars," in *Radar—Present and Future*, IEE Conf. Publ. no. 105, London, UK, October 1973.
46. N. Khoury and J. S. Hoyle, "Clutter maps: Design and performance," in *IEEE Nat. Radar Conf.*, Atlanta, GA, 1984.
47. G. V. Trunk et al., "False alarm control using doppler estimation," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-26, pp. 146–153, January 1990.
48. W. W. Shrader, inventor, "Sensitivity Velocity Control," U.S. Patent 5,134,410, July 1992.
49. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Information Theory*, vol. IT-20, no. 5, pp. 591–598, September 1974.
50. D. R. A. McMahon and R. F. Barrett, "An efficient method for the estimation of the frequency of a single tone in noise from the phases of discrete Fourier transforms," *Signal Processing*, vol. 11, pp. 169–177, 1986.
51. J. W. Taylor, "Receivers," Chap. 3 in *Radar Handbook*, 2nd Ed., M. I. Skolnik (ed.), New York: McGraw-Hill, 1990, pp. 323–325.
52. J. Y. N. Cho et al., "Range-velocity ambiguity mitigation schemes for the enhanced terminal doppler weather radar," in *31st Conference on Radar Meteorology*, Seattle, WA, 2003, pp. 463–466.
53. W. W. Shrader, "Radar technology applied to air traffic control," *IEEE Trans. Communications*, vol. 21, no. 5, pp. 591–605, May 1973.
54. W. W. Shrader, "MTI radar," Chap. 17 in *Radar Handbook*, M. I. Skolnik (ed.), New York: McGraw-Hill, 1970, pp. 17–56.

Chapter 3
Airborne MTI

James K. Day

Lockheed Martin Corporation

Fred M. Staudaher*

Naval Research Laboratory (retired)

3.1 SYSTEMS USING AIRBORNE MTI TECHNIQUES

Airborne search radars were initially developed for the detection of ships by long-range patrol aircraft. During the latter part of World War II, airborne early-warning (AEW) radars were developed by the U.S. Navy to detect low-flying aircraft approaching a task force below the radar coverage of the ship's antenna. The advantage of the airborne platform in extending the maximum detection range for air and surface targets is apparent when one considers that the radar horizon is 12 nmi for a 100-ft antenna mast compared with approximately 195 nmi for a 25,000-ft aircraft altitude.

The aircraft carrier–based E-2D aircraft (Figure 3.1) uses AEW radar as the primary sensor in its airborne tactical data system. These radars with their extensive field of view are required to detect small airborne targets against a background of sea and land clutter. Because their primary mission is to detect low-flying aircraft, they cannot elevate their antenna beam to eliminate the clutter. These considerations have led to the development of airborne MTI (AMTI)[1,2,3] radar systems similar to those used in surface radars,[1,4–6] discussed in the preceding chapter.

The mission requirements for an AEW radar drive the need for 360° azimuthal coverage and long-range detection capability. The 360° azimuthal coverage requirement is because the AEW radar system is generally required to provide the first detection of airborne targets, without any a priori knowledge of the location of these targets. AEW systems have generally been developed at lower frequencies—this can be understood by reviewing the surveillance radar range equation:

$$R_{\max} = \frac{P_a A_e \sigma_t}{(4\pi)\, kT_0 F_n L\, (S/N)_0} \frac{t_s}{\Omega} \tag{3.1}$$

* Sections 3.4 through 3.8 and 3.10 were taken primarily from the second edition of the *Radar Handbook*, Chapter 16, authored by Fred Staudaher, with revisions made by James Day. The remaining sections of the chapter were authored by James Day.

FIGURE 3.1 **E**-2D airborne early-warning (AEW) aircraft showing rotodome housing the antenna

where t_s is the scan time and Ω is the surveillance volume coverage requirement (product of the azimuth and elevation angles).

As long as the beamwidths of the radar (in azimuth and elevation) are smaller than the region to be surveilled, this equation is not directly dependent upon frequency. However, key parameters in this equation are dependent upon frequency. Particularly, propagation losses for low altitude targets and target RCS (for some target types) are generally advantageous for lower frequencies. The result is that AEW systems have been developed at UHF, L band, and S band frequencies.

Airborne MTI radar systems have also been utilized to acquire and track targets in interceptor fire control systems. In this application, the systems have to discriminate against clutter only in the vicinity of a prescribed target. This allows the system to be optimized at the range and angular sector where the target is located. MTI is also used to detect moving ground vehicles by reconnaissance and tactical fighter aircraft.

The environment of high platform altitude, mobility, and speed, coupled with restrictions on size, weight, and power consumption, present a unique set of problems to the designer of airborne MTI systems. This chapter will be devoted to considerations unique to the airborne environment.

3.2 COVERAGE CONSIDERATIONS

Search radars generally require 360° azimuthal coverage. This coverage is difficult to obtain on an aircraft since mounting an antenna in the clear presents major drag, stability, and structural problems. When extensive vertical coverage is required, the aircraft's planform and vertical stabilizer distort and shadow the antenna pattern. Analysis of tactical requirements may show that only a limited coverage sector is required. However, this sector usually has to be capable of being positioned over the full 360° relative to the aircraft's heading because of the requirements for coverage

FIGURE 3.2 Boeing 737-700 Wedgetail aircraft showing antennas mounted above the fuselage

while reversing course, large crab angles when high winds are encountered, the need to position ground track in relation to wind, nontypical operating situations, and operations requirements for coverage while proceeding to and from the station.

However, in the 1990s and 2000s, a number of systems have been developed that provide phased array performance in an airborne platform. The Multi-Role Electronically Scanned Array (MESA) radar developed by Northrop Grumman on a Boeing 737-700 for the Australian Wedgetail program is an example (see Figure 3.2). An alternate solution that combines mechanical scanning in conjunction with electronic scanning is in development with the AN/APY-9 radar for the E-2D aircraft (follow-up to the U.S. Navy's E-2C aircraft).

3.3 AIRBORNE MTI PERFORMANCE DRIVERS

The performance of airborne MTI systems are primarily determined by motion effects induced on the clutter echoes (platform motion, antenna scanning motion, and clutter internal motion), the processing techniques used to enhance target detection and maximize clutter cancellation, and the hardware stability limitations of the radar. This chapter will discuss the motion effects as well as the performance of various processing techniques.

3.4 PLATFORM MOTION AND ALTITUDE EFFECTS ON MTI PERFORMANCE

MTI discriminates between airborne moving targets and stationary land or sea clutter. However, in the airborne case, the clutter moves with respect to the moving airborne platform. It is possible to compensate for the mean clutter radial velocity by using

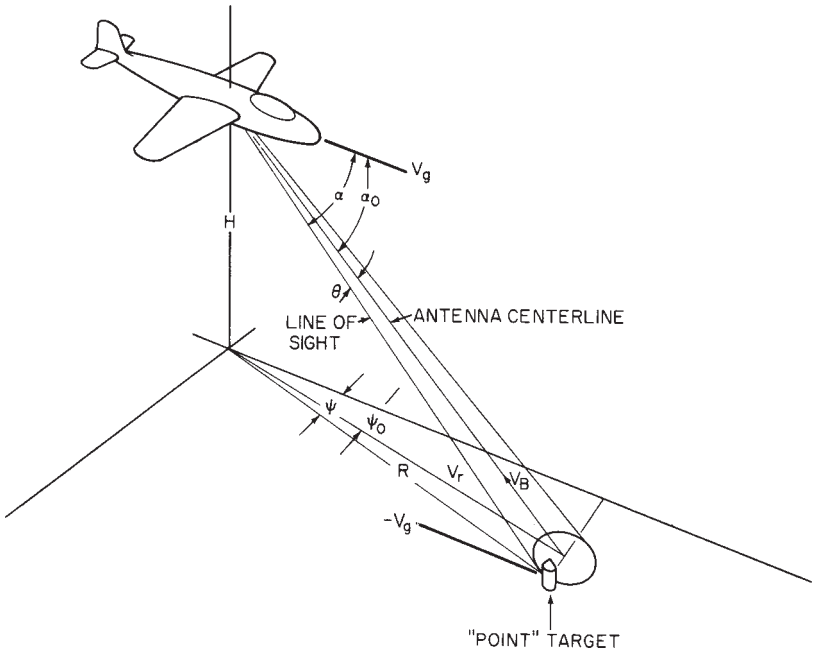


FIGURE 3.3 Defining geometry: α_0 = antenna pointing angle; α = line-of-sight angle; θ = angle from antenna centerline; V_g = aircraft ground speed; V_r = radial velocity of point target; V_B = radial velocity along antenna centerline (boresight); ψ_0 = antenna azimuth angle; ψ = azimuth angle; R = ground range to point target; and H = aircraft height

techniques such as time-averaged-clutter coherent airborne radar (TACCAR). This technique attempts to center the largest return from main-beam clutter at zero doppler frequency such that a simple MTI filter, also centered at zero doppler frequency, will cancel the main-beam clutter.

As shown in Figure 3.3, the apparent radial velocity of the clutter is $V_r = -V_g \cos \alpha$, where V_g is the ground speed of the platform and α is the angle subtended between the line-of-sight to a point on the Earth's surface and the aircraft's velocity vector. Figure 3.4 shows the loci of constant radial velocity along the surface. In order to normalize the figure, a flat earth is assumed, and the normalized radial velocity $V_n = V_r/V_g$ is presented as a function of azimuth angle ψ and normalized ground range R/H, where H is the aircraft's altitude.

Instead of a single clutter doppler frequency corresponding to a constant radial velocity (V_B in Figure 3.3) determined by the antenna pointing angle α_0, the radial sees a continuum of velocities. This results in a frequency spectrum at a particular range whose shape is determined by the antenna pattern that intersects the surface, the reflectivity of the clutter, and the velocity distribution within the beam. Furthermore, since V_r varies as a function of range at a particular azimuth ψ, the center frequency and spectrum shape vary as a function of range and azimuth angle ψ_0.

When the antenna is pointing ahead, the predominant effect is the variation of the center frequency corresponding to the change in α_0 with range. When the antenna is pointing

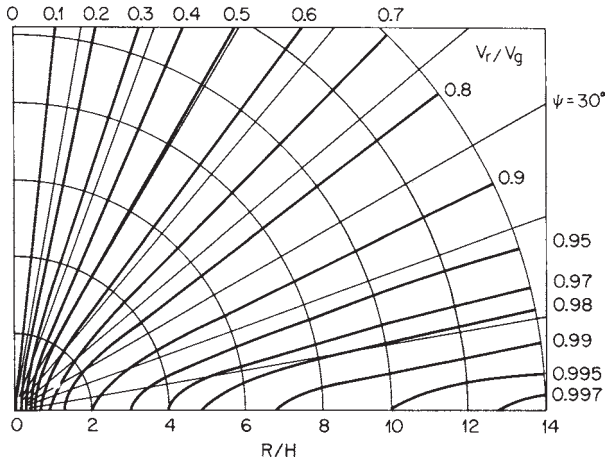


FIGURE 3.4 Loci of constant normalized radial velocity V_r/V_g as a function of aircraft range-to-height ratio R/H and azimuth angle ψ

abeam, the predominant effect is the velocity spread across the antenna beamwidth. These are classified as the slant-range effect and the platform-motion effect, respectively.

Effect of Slant Range on Doppler Offset. The antenna boresight velocity V_B is the ground-velocity component along the antenna centerline (boresight) and is given as $(-V_g \cos \alpha_0)$. If the clutter surface were coplanar with the aircraft, this component would be equal to $(-V_g \cos \psi_0)$ and would be independent of range. The ratio of the actual boresight velocity to the coplanar boresight velocity is defined as the normalized boresight-velocity ratio:

$$VBR = \frac{\cos \alpha_0}{\cos \psi_0} = \cos \phi_0 \tag{3.2}$$

where ϕ_0 is the depression angle of the antenna centerline from the horizontal. Figure 3.5 shows the variation of the normalized boresight-velocity ratio as a function of slant range for a curved earth and different aircraft altitudes. The variation is fairly rapid for slant ranges less than 15 nmi.

It is desirable to center the clutter spectrum in the notch (i.e., minimum-response region) of the AMTI filter in order to obtain maximum clutter rejection. This can be accomplished by offsetting the IF or RF frequency of the radar signal by an amount equal to the average doppler frequency of the clutter spectrum. Because the clutter center frequency varies with range and azimuth when the radar is moving, it is necessary for the filter notch to track the doppler-offset frequency, using an open- or closed-loop control system such as TACCAR, described below.

An example of a received clutter spectrum given an antenna response is shown in Figure 3.6*a*. The TACCAR frequency offset then shifts main-beam clutter to zero doppler, as shown in Figure 3.6*b*.

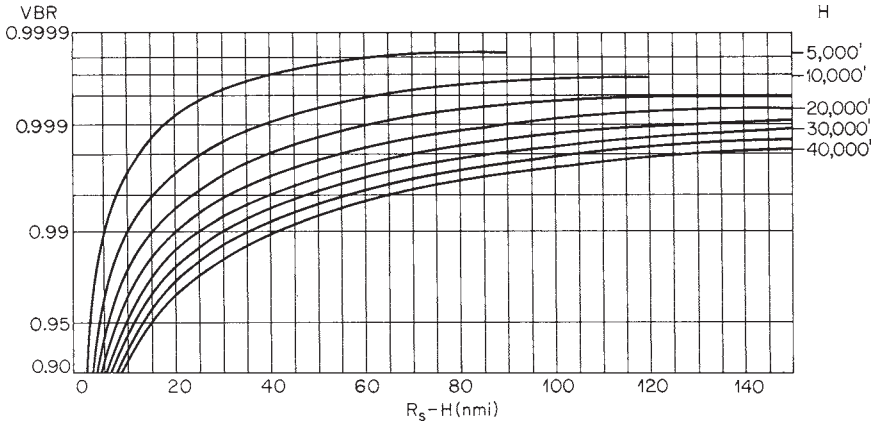


FIGURE 3.5 Normalized boresight-velocity ratio *VBR* as a function of the difference between slant range R_s and aircraft altitude H for different aircraft altitudes

TACCAR. The MIT Lincoln Laboratory originally developed TACCAR to solve the AMTI radar problem. The requirements and thus the implementation of TACCAR change depending upon the type of clutter cancellation processing employed. After many other approaches, it was recognized that if one used the clutter return rather than the transmit pulse to phase-lock the radar to the clutter filter, one could center the clutter in the filter stopband. The clutter phase varies from range cell to range cell owing to the distribution of the location of the scatterers in azimuth. Hence, it is necessary to average the return for as long an interval as possible. TACCAR is used to describe the centering of the returned clutter spectrum to the zero filter frequency. Since the technique compensates for drift in the various system elements and biases in the mean doppler frequency due to ocean currents, chaff, or weather clutter, it is used in shipboard and land-based radars as well as airborne radar.

A functional block diagram of an airborne radar employing TACCAR is shown in Figure 3.7. The clutter error signal is obtained by measuring the pulse-to-pulse phase shift $\omega_d T_p$ of the clutter return. This provides a very sensitive error signal. The averaged error signal controls a voltage-controlled coherent master oscillator (COMO), which determines the transmitted frequency of the radar. The COMO is slaved to

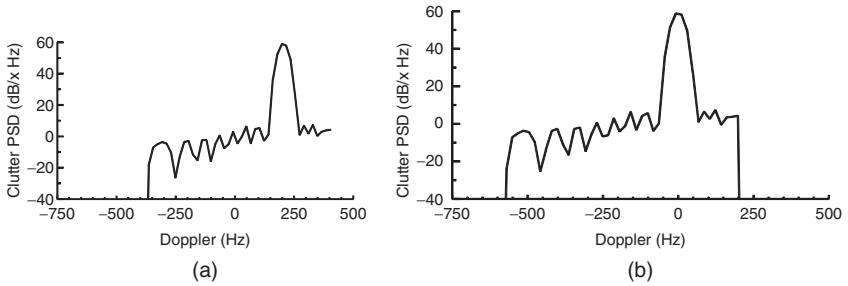


FIGURE 3.6 Clutter Power Spectral Density (PSD) response through antenna pattern: *a*) without TACCAR frequency offset and *b*) with TACCAR frequency offset

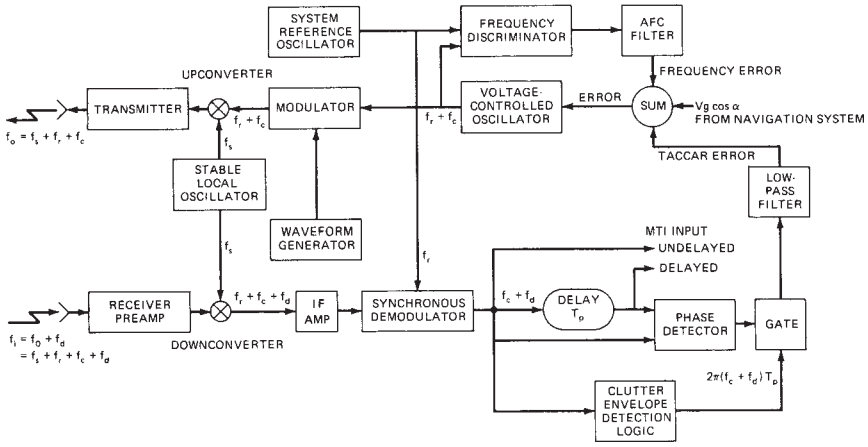


FIGURE 3.7 Block diagram of a radar illustrating the signal flow path of the TACCAR control loop

the system reference oscillator frequency via the automatic frequency control (AFC) loop shown in Figure 3.7. This provides a stable reference in the absence of clutter. An input from the aircraft inertial navigation system and the antenna servo provide a predicted doppler offset. These inputs allow the TACCAR system to provide a narrow-bandwidth correction signal.

Because of the noisy nature of the clutter signal, the need to have the control system bridge regions of weak clutter return, and the requirement not to respond to the doppler shift of a true target, the control system usually tracks the azimuth variation of a specific radar range interval. The maximum range of this interval is chosen so that clutter will be the dominant signal within the interval. The minimum range is chosen to exclude signals whose average frequency differs substantially from the frequency in the region of interest.

Alternate approaches to providing this frequency offset can be implemented with digital exciters or on receive. For some applications, it may be necessary to use multiple control loops, each one covering a specific range interval, or to vary the offset frequency in range. This is possible if the frequency offset is implemented on receive (but not on transmit). At any particular range, the filter notch is effectively at one frequency and the center frequency of the clutter spectrum at another. The difference between these frequencies results in a doppler-offset error, as shown in Figure 3.8. The clutter spectrum will extend into more of the filter passband, and the clutter improvement factor will be degraded. The required accuracy for the TACCAR control loop can be relaxed if the MTI filter is an adaptive filter, such as with space-time adaptive processing (discussed later in this chapter). This is because the adaptive filter will adjust to the received signals and optimize clutter cancellation.

Without adaptive adjustment, Figure 3.9 shows the improvement factor for single- and double-delay cancelers as a function of the ratio of the notch-offset error to the pulse repetition frequency (PRF) for different clutter spectral widths. Fortunately, the platform-motion spectrum is narrow in the forward sector of coverage where offset error is maximum. An offset error of one-hundredth of the PRF would yield a 26 dB improvement factor for a double canceler with an input clutter spectrum whose width

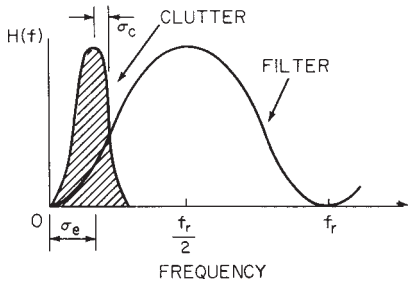


FIGURE 3.8 Effect of doppler-offset error; f_r = PRF

was 3% of the PRF. If the radar frequency were 10 GHz, PRF 1 kHz, and ground speed 580 kt, the notch would have to be held within 0.29 kt or $0.005V_g$.

Because of these requirements and the width of the platform-motion spectrum, stagger PRF systems must be chosen primarily on the basis of maintaining the stopband rather than flattening the passband. Similarly, higher-order delay-line filters (with or without feedback) are synthesized on the basis of stopband rejection. The limiting case is the narrowband filter bank where each individual filter consists of a small passband, the balance being stopband.

Improvement factor is an important metric, but in addition to this average metric defined across all doppler frequencies, it is often important to characterize the performance as a function of doppler frequency, particularly with coherent doppler filtering imbedded in the processing chain. With performance characterized versus doppler

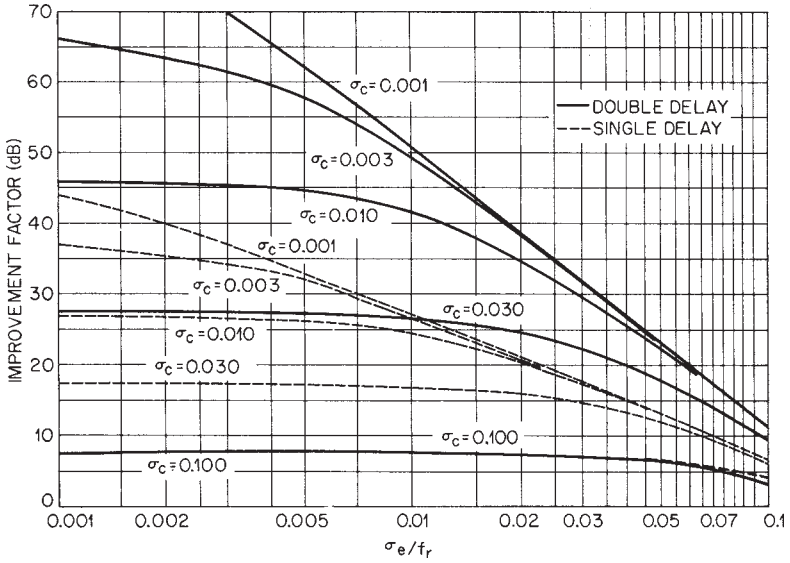


FIGURE 3.9 Improvement factor I versus normalized doppler offset σ_e as a function of clutter spectrum width σ_c

frequency, the radar design can then be evaluated through the complete detection chain and optimized in conjunction with any multiple PRF stagger waveforms utilized to bridge MTI blind regions.

Platform-Motion Effect. To an airborne radar, a clutter scatterer appears to have a radial velocity that differs from the antenna-boresight radial velocity at the same range by

$$\begin{aligned} V_e &= V_r - V_B \\ &= V_g \cos\alpha_0 - V_g \cos\alpha \\ &= V_g[\cos\alpha_0 - \cos(\alpha_0 + \theta)] \\ &= V_x \sin\theta + 2V_y \sin^2\frac{\theta}{2} \end{aligned} \tag{3.3}$$

for small values of θ and depression angle ϕ_0, where V_x is the horizontal component of velocity perpendicular to the antenna boresight and V_y is the component along the antenna boresight. θ is the azimuthal angle from the antenna boresight, or the intersection of the vertical plane containing the boresight with the ground. The corresponding doppler frequency, when α_0 is a few beamwidths from ground track, is

$$f_d = \frac{2V_x}{\lambda}\sin\theta \approx \frac{2V_x}{\lambda}\theta \tag{3.4}$$

This phenomenon results in a platform-motion clutter power spectrum that is weighted by the antenna's two-way power pattern in azimuth. The true spectrum may be approximated by a gaussian spectrum,

$$H(f) = e^{-\frac{1}{2}\left(f_d/\sigma_{\text{pm}}\right)^2} = e^{-2\left(V_x\theta/\lambda\sigma_{\text{pm}}\right)^2} \approx G^4(\theta) \tag{3.5}$$

$G^4(\theta)$, the two-way power pattern of the antenna, is 0.25 when $\theta = \theta_a/2$, where θ_a is the half-power beamwidth, which can be approximated by λ/a, a being the effective horizontal aperture width. Thus,

$$e^{-\frac{1}{2}\left(V_x/a\sigma_{\text{pm}}\right)^2} = 0.25$$

or

$$\sigma_{\text{pm}} = 0.6\frac{V_x}{a} \tag{3.6}$$

where V_x and a are in consistent units. This value is lower than ones derived by other authors.[4,5] However, it agrees with more exact analysis of antenna radiation patterns and experimental data analyzed by F. Staudaher.

A more exact value of the parameter σ_{pm} may be obtained by matching a two-way power pattern of interest with the gaussian approximation at a specific point on the pattern, determining the standard deviation of θ by using statistical techniques or fitting

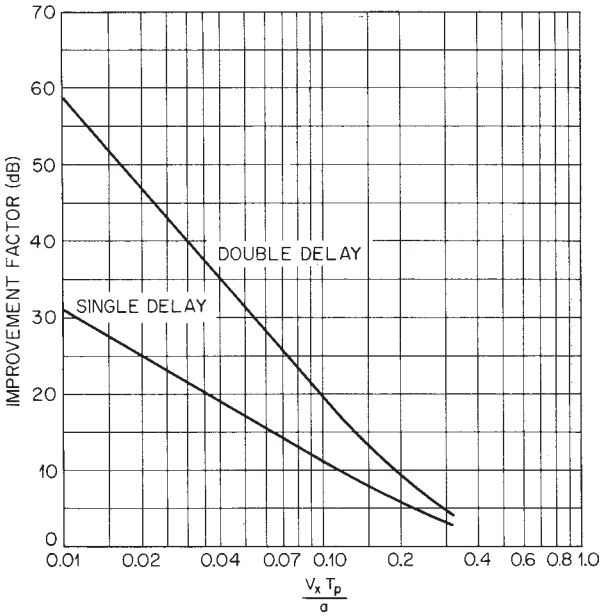


FIGURE 3.10 Effect of platform motion on the MTI improvement factor as a function of the fraction of the horizontal antenna aperture displaced per interpulse period, $V_x T_p/a$

the pattern and using numerical methods. The calculation of the improvement factor can be performed by averaging the resultant residue power, obtained by summing the signal phasors at specific values of θ, from null to null of the antenna pattern.

Figure 3.10 shows the effect of platform motion on the MTI improvement factor as a function of the aperture displaced in the plane of the aperture per interpulse period T_p. A 5.4% displacement reduces the double-delay improvement factor to 30 dB. This corresponds to a speed of 325 kt if the system has a PRF of 1000 Hz and a 10-ft antenna aperture. For a single-delay system, the displacement has to be held to 1.1% for a 30 dB performance limit.

3.5 PLATFORM-MOTION COMPENSATION ABEAM

The deleterious effects of platform motion can be reduced by physically or electronically displacing the antenna phase center along the plane of the aperture. This is referred to as the *displaced phase center antenna (DPCA) technique*.[7–11] In addition, some forms of space-time adaptive processing are expressly developed to improve clutter cancellation with an adaptive filter, electronically displacing the antenna phase center.

Electronically Displaced Phase Center Antenna. Figure 3.11*a* shows the pulse-to-pulse phase advance of an elemental scatterer as seen by the radar receiver.

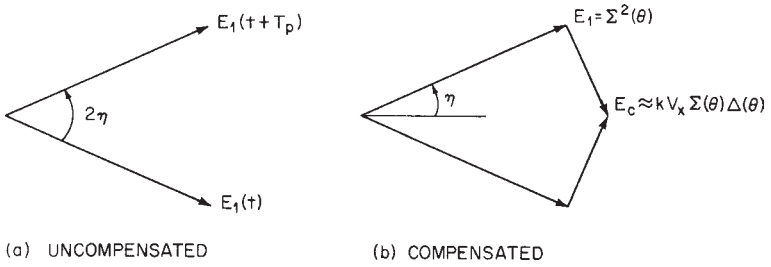


FIGURE 3.11 Phasor diagram showing the return from a point scatterer due to platform motion

The amplitude E_1 of the received signal is proportional to the two-way antenna field intensity. The phase advance is

$$2\eta = 2\pi f_d T_p = \frac{4\pi V_x T_p \sin\theta}{\lambda} \tag{3.7}$$

where f_d = doppler shift of scatterer (Eq. 3.4)
T_p = interpulse period

Figure 3.11*b* shows a method of correcting for the phase advance η. An idealized correction signal E_c is applied, leading the received signal by 90° and lagging the next received signal by 90°. For exact compensation, the following relation would hold:

$$E_c = E_1 \tan\eta = \Sigma^2(\theta)\tan\frac{2\pi V_x T_p \sin\theta}{\lambda} \tag{3.8}$$

This assumes a two-lobe antenna pattern similar to that in a monopulse tracking radar. Two receivers are used, one supplying a sum signal, $\Sigma(\theta)$, and the other a difference signal, $\Delta(\theta)$. The difference signal is used to compensate for the effects of platform motion.

If the system is designed to transmit the sum pattern $\Sigma(\theta)$ and receive both $\Sigma(\theta)$ and a difference pattern $\Delta(\theta)$, then at the design speed the received signal $\Sigma(\theta)\Delta(\theta)$ can be applied as the correction signal. The actual correction signal used to approximate E_c is $k\,\Sigma(\theta)\Delta(\theta)$, where k is the ratio of the amplification in the sum and difference channels of the receiver.

A uniformly illuminated monopulse array[12] has the difference signal Δ in quadrature with the sum and has the amplitude relationship

$$\Delta(\theta) = \Sigma(\theta)\tan\left(\frac{\pi W}{\lambda}\sin\theta\right) \tag{3.9}$$

where W is the distance between the phase centers of the two halves of the antenna. Hence, a choice of $W = 2V_xT_p$ and $k = 1$ would ideally result in perfect cancellation.

In practice, a sum pattern is chosen based on the desired beamwidth, gain, and sidelobes for the detection system requirements. Then the difference pattern $\Delta(\theta)$ is synthesized independently, based on the relationship required at design radar platform

speed and allowable sidelobes. The two patterns may be realized by combining the elements in separate corporate-feed structures.

Figure 3.12 shows the idealized improvement factor as a function of normalized aperture movement for a double-delay canceler. The improvement factor shown is the improvement factor for a point scatterer averaged over the null-to-null antenna beamwidth. In one case, the gain ratio k is optimized at each value of pulse-to-pulse displacement. In the other compensated case, the optimum gain ratio k is approximated by the linear function of interpulse platform motion kV_x.

A block diagram of the double-delay system is shown in Figure 3.13. A single-delay system would not have the second delay line and subtractor. The normally required circuitry for maintaining coherence, gain and phase balance, and timing is not shown. The speed control V_x is bipolar and must be capable of reversing the sign of the $\Delta(\theta)$ signal in each channel when the antenna pointing angle changes from the port to the starboard side of the aircraft.

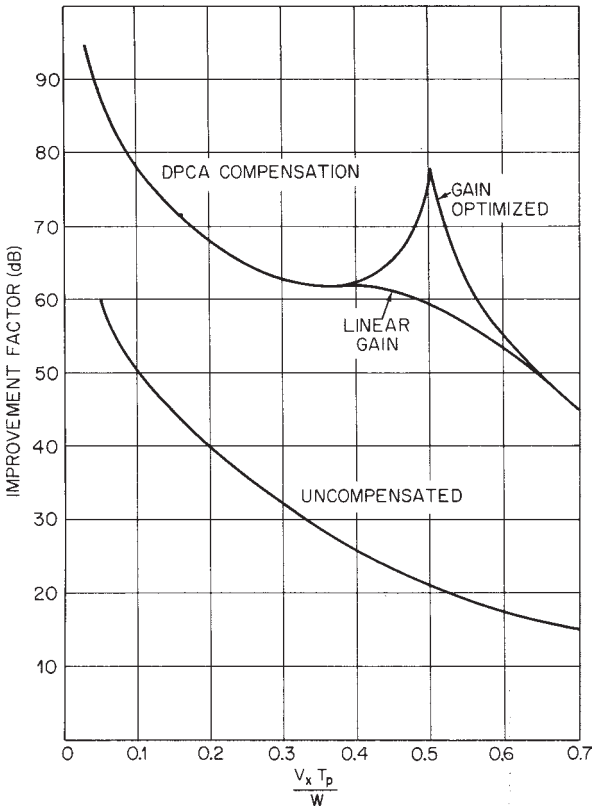


FIGURE 3.12 MTI improvement factor for DPCA compensation as a function of the fraction of the horizontal phase center separation W that the horizontal antenna aperture is displaced per interpulse period, V_xT_p/W. $W = 0.172a$, where a is the horizontal aperture length.

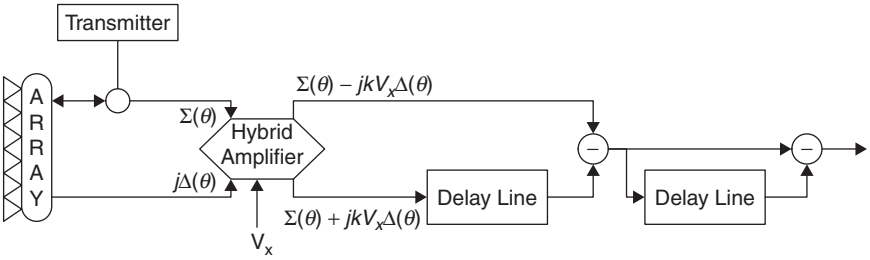


FIGURE 3.13 Simplified double-delay DPCA mechanization

The hybrid amplifier shown has two input terminals that receive $\Sigma(\theta)$ and $j\Delta(\theta)$ and amplify the $\Delta(\theta)$ channel by kV_x relative to the $\Sigma(\theta)$ channel. The output terminals produce the sum and difference of the two amplified input signals. Because DPCA compensates for the complex signal, both amplitude and phase information must be retained. Therefore, these operations usually occur at RF or IF. Digital compensation can be used if synchronous detection and analog to digital (A/D) conversion are performed and the components are treated as complex phasors. Furthermore, the operations must be linear until the sum signal and difference signals have been processed by the hybrid amplifier. After this single-pulse combination, the actual double cancellation can be performed by any conventional MTI processing techniques.

Power in the Antenna Sidelobes. Airborne systems are limited in their ability to reject clutter due to the power returned by the antenna sidelobes. The full 360° azimuthal pattern sees velocities from $-V_g$ to $+V_g$. The compensation circuits offset the velocity by an amount corresponding to the antenna boresight velocity V_B, but the total range of doppler frequencies corresponding to $2V_g$ is obtained because of echoes received via the sidelobes. For airborne systems with low PRFs, these doppler frequencies can cover several multiples of the PRF so that the sidelobe power is folded into the filter. This limitation is a function of the antenna pointing angle, the MTI filter response, and the sidelobe pattern. If the sidelobes are relatively well distributed in azimuth, a measure of performance can be obtained by averaging the power returned by the sidelobes.

The limiting improvement factor due to sidelobes is

$$I_{\text{sl limit}} = \frac{K\int_{-\pi}^{\pi} G^4(\theta)\,d\theta}{\int_{\text{sl}} G^4(\theta)\,d\theta} \tag{3.10}$$

where the lower integral is taken outside the main-beam region. Main-beam effects would be included in the platform-motion improvement factor. The constant K is the noise normalization factor for the MTI filter ($K = 2$ for single delay and 6 for double delay.) $G^4(\theta)$ is the two-way power of the antenna in the plane of the ground surface.

The DPCA performance described in the preceding subsection can be analyzed on the basis of radiation patterns or the equivalent aperture distribution function.[8] If the radiation pattern is used, the composite performance may be obtained either by applying the pattern functions over the entire 360° pattern or by combining the improvement

factors for the DPCA main-beam and the sidelobe regions in the same manner as parallel impedances are combined:

$$\frac{1}{I_{\text{total}}} = \frac{1}{I_{sl}} + \frac{1}{I_{\text{DPCA}}} \tag{3.11}$$

If the aperture distribution is used, the sidelobe effects are inherent in the analysis. Care must be taken, however—if the array or reflector function is used without considering the weighting of the elemental pattern or the feed distribution, the inherent sidelobe pattern can obscure the main-beam compensation results.

Again, the performance versus doppler frequency is important for evaluating overall radar detection performance. Antenna sidelobe limited performance can be approximated by performing the lower integral of Eq. 3.10 over those angles that map into a given doppler filter's passband. The noise normalization term, k, must also be modified to reflect the cascaded noise gain of the MTI and doppler filter bank as

$$N_g(k) = 6\sum_{i=1}^{N} W_i^2 - 8\sum_{i=1}^{N-1} W_i W_{i+1}\cos(2\pi k/N) + 2\sum_{i=1}^{N-2} W_i W_{i+2}\cos(4\pi k/N); \quad k = 0, N-1 \tag{3.12}$$

for three-pulse MTI and cascaded N-pulse doppler filter bank, where W_i are the doppler filter weights, or

$$N_g(k) = 2\sum_{i=1}^{N} W_i^2 - 2\sum_{i=1}^{N-1} W_i W_{i+1}\cos(\pi k/N); \quad k = 0, N-1 \tag{3.13}$$

for two-pulse MTI and cascaded N-pulse doppler filter bank.

3.6 SCANNING-MOTION COMPENSATION

Figure 3.14*a* shows a typical antenna main-beam radiation pattern and the response of a point scatterer for two successive pulses when the antenna is scanning. It is seen that the signals returned would differ by $\Delta G^2(\theta)$. This results in imperfect cancellation due to scanning. The average effect on the improvement factor can be obtained by integrating this differential effect $\Delta G^2(\theta)$ over the main beams:

$$I_{\text{scan}} = \frac{2\int_{-\theta_0}^{\theta_0} |G(\theta)|^2 \, d\theta}{\int_{-\theta_0}^{\theta_0} |G(\theta + T_p\dot{\theta}) - G(\theta)|^2 \, d\theta} \quad \text{for single-delay cancellation} \tag{3.14a}$$

$$I_{\text{scan}} = \frac{6\int_{-\theta_0}^{\theta_0} |G(\theta)|^2 \, d\theta}{\int_{-\theta_0}^{\theta_0} |G(\theta + T_p\dot{\theta}) - 2G(\theta) + G(\theta - T_p\dot{\theta})|^2 \, d\theta} \quad \text{for double-delay cancellation} \tag{3.14b}$$

where θ_0 = null of main beam
$G(\theta)$ = two-way voltage pattern

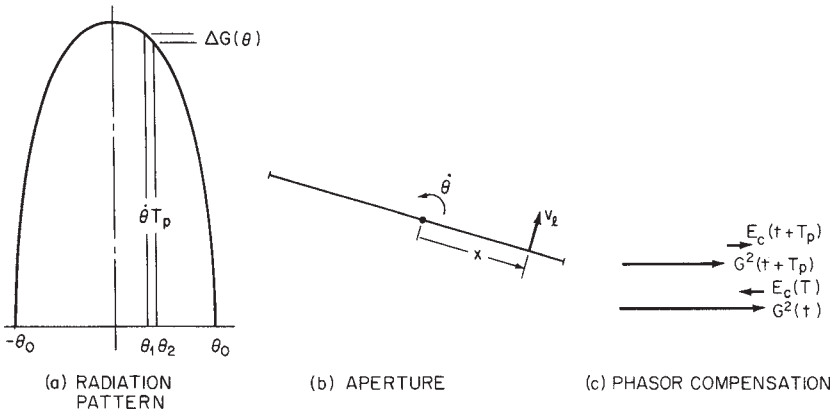


FIGURE 3.14 Antenna scanning effects: (*a*) as seen by the antenna radiation pattern, due to the apparent change in azimuth of the scatterer, $\theta_2 - \theta_1 = \dot{\theta}\, T_p$; (*b*) as seen by the aperture illumination function, due to the apparent motion, $v_1 = x\dot{\theta}$, of the scatterer relative to the antenna at position x; and (*c*) step-scan compensation of two received phasors

In order to treat scanning motion in the frequency domain, the apparent clutter velocity seen by the scanning antenna is examined to determine the doppler frequency. Each element of an array or incremental section of a continuous aperture can be considered as receiving a doppler-shifted signal due to the relative motion of the clutter. The power received by the element is proportional to the two-way aperture power distribution function $F_2(x)$ at the element.

In addition to the velocity seen by all elements because of the motion of the platform, each element sees an apparent clutter velocity due to its rotational motion, as illustrated in Figure 3.14*b*. The apparent velocity varies linearly along the aperture. Hence, the two-way aperture distribution is mapped into the frequency domain. The resulting power spectrum due to the antenna scanning is

$$H(f) = F_2\left(\frac{\lambda f}{2\dot{\theta}}\right) \qquad 0 \le f \le \frac{a\dot{\theta}}{\lambda} \tag{3.15}$$

where $\dot{\theta}$ = antenna rotation rate
a = horizontal antenna aperture

This spectrum can be approximated by a gaussian distribution with standard deviation

$$\sigma_c = 0.265\frac{f_r}{n} = 0.265\frac{\dot{\theta}}{\theta_a} \approx 0.265\frac{a\dot{\theta}}{\lambda} \tag{3.16}$$

where λ and a are in the same units, θ_a is the one-way half-power beamwidth, and n is the number of hits per beamwidth. The approximation $\theta_a \approx \lambda/a$ is representative of an antenna distribution yielding acceptable sidelobe levels.

It can be seen that the antenna pattern pulse-to-pulse differential gain is

$$\Delta G^2(\theta) = \frac{dG^2(\theta)}{d\theta}\Delta\theta = \frac{dG^2(\theta)}{d\theta}\dot{\theta}T_p \tag{3.17}$$

This suggests[7,13] that a correction signal in the reverse sense to $\Delta G^2(\theta)$ be applied, as shown in Figure 3.14*c*. Half the correction is added to one pulse and half subtracted from the other, so that

$$\begin{aligned}\text{Correction signal} &= \frac{\Delta G^2(\theta)}{2} = \frac{\dot{\theta} T_p}{2} \frac{d\Sigma^2(\theta)}{d\theta} \\ &= \dot{\theta} T_p \Sigma(\theta) \frac{d\Sigma(\theta)}{d\theta}\end{aligned} \tag{3.18}$$

where $\Sigma^2(\theta)$ was substituted for $G^2(\theta)$. The radar transmits a sum pattern $\Sigma(\theta)$ and receives on the difference pattern $\Delta(\theta)$, so that the received signal is proportional to the product of the two. If the signal received on the difference pattern is used as the correction, we have

$$E_c = \Delta(\theta)\Sigma(\theta) \tag{3.19}$$

By comparing Eqs. 3.18 and 3.19, we see that for E_c to approximate the correction signal, the difference patterns should be

$$\Delta(\theta) = \dot{\theta} T_p \frac{d\Sigma(\theta)}{d\theta} \tag{3.20}$$

The derivative of the sum pattern is similar to a difference pattern in that it is positive at the main-beam null, $-\theta_0$, decreases to zero on the antenna centerline, and then goes negative until θ_0.

Referring to Figure 3.13, one observes that the mechanization for scan compensation is fundamentally similar to the DPCA mechanization except that the difference signal is applied in phase with the sum signal and amplified by an amount determined by the antenna rotation per interpulse period.

The signals required, if the transmission signal $\Sigma(\theta)$ that appears in each channel is neglected, are $\Sigma(\theta) \pm l\dot{\theta} T_p \Delta(\theta)$, where l is the ratio of the amplification in the two channels chosen to maximize the clutter rejection. The required difference-pattern slope is determined by the derivative of the scan pattern, which differs from the DPCA criterion. This technique is known as *step-scan compensation* because the system electronically points the antenna slightly ahead of and behind of boresight each pulse so that a leading and lagging pair are taken from successive returns to obtain the effect of the antenna remaining stationary.

Figure 3.15 shows the improvement obtained by Dickey and Santa[7] for single-delay cancellation.

Compensation-Pattern Selection. Selection of the compensation pattern depends on the level of system performance required, the type of MTI filtering used, the platform velocity, scan rate, and the characteristics required by normal radar parameters such as resolution, distortion, gain, sidelobes, etc. For instance, an exponential pattern and its corresponding difference pattern are excellent for single-delay-cancellation DPCA but are unsatisfactory when double-delay cancellation is used. This is because the single-delay canceler requires the best match between the actual pattern and the required pattern near boresight, whereas double cancellation requires the best match

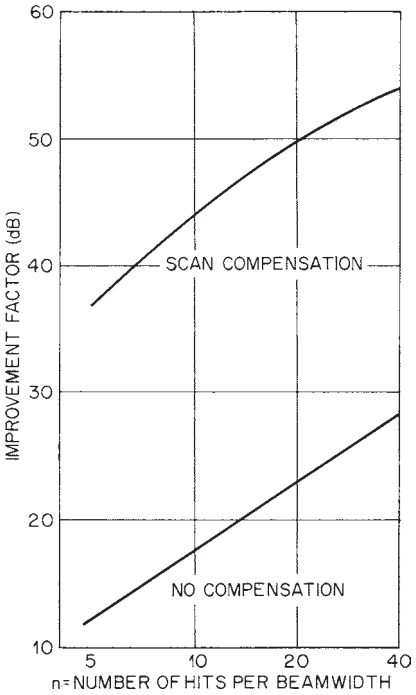


FIGURE 3.15 MTI improvement factor for a step-scan compensation of a single-delay canceler as a function of the number of hits per beamwidth. The antenna pattern is $(\sin x)/x$.

on the beam shoulder. Step-scan compensation usually requires the difference-pattern peaks to be near the nulls of the sum pattern to match.

Grissetti et al.[13] have shown that for step-scan compensation the improvement factor for single-delay cancellation increases as a function of the number of hits at 20 dB/decade; for the first-derivative*-type step-scan compensation, at the rate of 40 dB/decade; and with first- and second-derivative compensation, at the rate of 60 dB/decade. Hence, for a ground-based system that is limited by scan rate, one should improve the compensation pattern rather than use a higher-order MTI canceler. However, airborne systems are primarily limited by platform motion and require both better cancelers and compensation for operation in a land-clutter environment. In the sea-clutter environment, the system is usually dominated by the spectral width of the velocity spectrum or platform motion rather than scanning. The applicability of DPCA or step-scan compensation in the latter case is dependent on the particular system parameters.

* The compensation required by $\Delta G^2(\theta)/2$ can be determined from a Taylor's series expansion of $G^2(\theta)$. In the preceding discussion, we used the first derivative. Using higher-order terms gives an improved correction signal.

3.7 SIMULTANEOUS PLATFORM MOTION AND SCAN COMPENSATION

In AMTI systems having many hits per scan, scanning is a secondary limitation for an uncompensated double canceler. However, the performance of a DPCA system is significantly reduced when it is scanned. This is due to the scanning modulation on the difference pattern used for platform-motion compensation.

Since the DPCA applies the difference pattern in quadrature to the sum pattern to compensate for phase error and step scan applies the difference pattern in phase to compensate for amplitude error, it is possible to combine the two techniques by properly scaling and applying the difference pattern both in phase and in quadrature. The scaling factors are chosen to maximize the improvement factor under conditions of scanning and platform motion.

The relationships for a double-delay (three-pulse) AMTI are shown in the phasor diagram in Figure 3.16. The phase advance between the first pair of pulses (first and second pulse for the three-pulse MTI) received by the sum pattern Σ is

$$2\eta_1 = \frac{4\pi T_p}{\lambda}\left[V_x\left(\sin\theta_2 - \sin\frac{\omega_r T_p}{2}\right) + V_y\left(\cos\frac{\omega_r Tp}{2} - \cos\theta_2\right)\right] \tag{3.21}$$

and the phase advance between the second pair of pulses (second and third pulse for the three-pulse MTI) is

$$2\eta_2 = \frac{4\pi T_p}{\lambda}\left[V_x\left(\sin\theta_2 + \sin\frac{\omega_r T_p}{2}\right) + V_y\left(\cos\frac{\omega_r Tp}{2} - \cos\theta_2\right)\right] \tag{3.22}$$

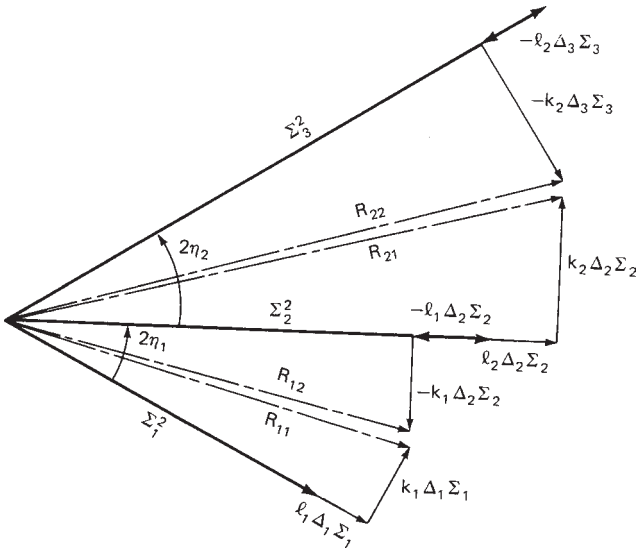


FIGURE 3.16 Phasor diagram for simultaneous scanning and motion compensation

where θ_2 is the direction of the clutter cell with respect to the antenna pointing angle when the second pulse is received and ω_r is the antenna scan rate. The subscripts on the received signals Σ_i and Δ_i indicate the pulse reception sequence.

The difference pattern Δ is used to generate an in-phase correction for scanning motion and a quadrature correction for platform motion. This process yields the set of resultant signals R_{ij}, where the subscript i denotes the pulse pair and the subscript j denotes the component of the pair. Because η_1 does not equal η_2, different weighting constants are required for each pulse pair. The values of k_1 for the quadrature correction of the first pulse pair, k_2 for the quadrature correction for the second pulse pair, l_1 for the in-phase correction for the first pulse pair, and l_2 for the second pulse pair are optimized by minimizing the integrated residue power over the significant portion of the antenna pattern, usually chosen between the first nulls of the main beam.

Figure 3.17 shows the sum and difference main-beam patterns for an aperture 20 wavelengths long. Figure 3.18 shows the residue for the case when the fraction of the horizontal aperture width a traveled per interpulse period T_p, $V_n = V_x T_p/a$, is equal to 0.04 and when the number of wavelengths that the aperture tip rotates per interpulse period, $W_n = a\omega_r T_p/2\lambda$, is equal to 0.04. The corresponding improvement factor is 52 dB.

The improvement factor is shown in Figure 3.19 for a range of normalized platform motion V_n as a function of normalized scanning displacements W_n. The nonscanning case is shown as $W_n = 0$. The improvement factors were computed for the 20-wavelength aperture patterns shown in Figure 3.17.

Andrews[14] has developed an optimization procedure for platform-motion compensation that rotates the phasors directly rather than by using a quadrature correction. The procedure determines the antenna feed coefficients for two compensation patterns, one of which, $C_1(\theta)$, is added to the sum pattern $\Sigma(\theta)$ and fed to the undelayed canceler

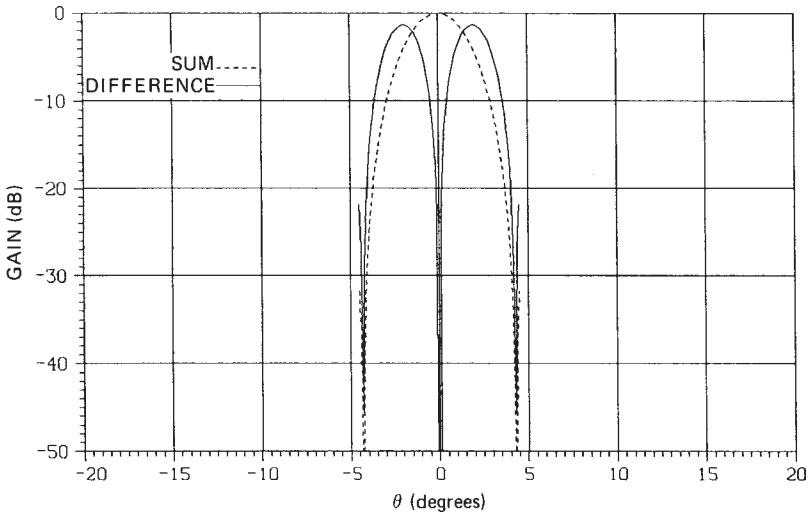


FIGURE 3.17 Sum and difference patterns used to determine DPCA performance

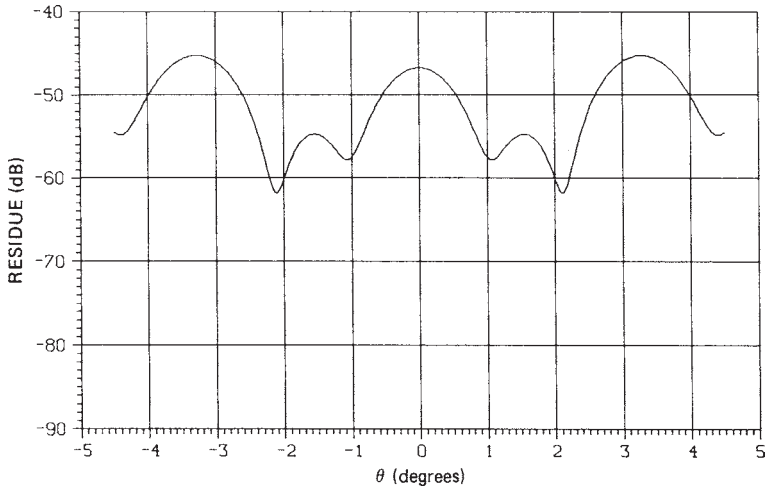


FIGURE 3.18 DPCA clutter residue versus angle for normalized displacement, $V_n = 0.04$, and normalized scanning motion, $W_n = 0.04$

path, and the other, $C_2(\theta)$, which is added to the sum pattern and fed to the delayed path as shown in Figure 3.20. The procedure was developed for a single-delay canceler and a nonscanning antenna. Andrews used the procedure to minimize the residue power over the full antenna pattern, which includes the main-beam and sidelobe regions.

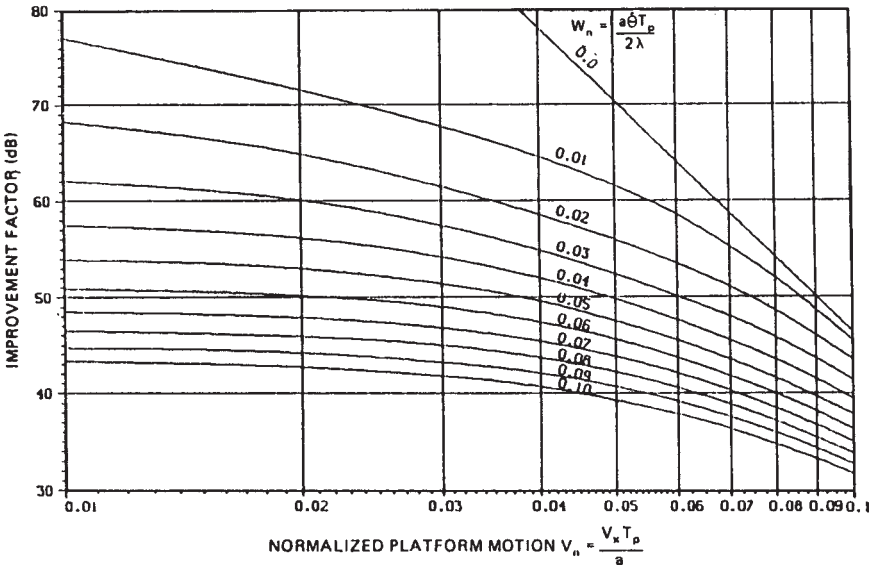


FIGURE 3.19 DPCA improvement factor versus normalized platform motion, V_n, as a function of normalized scanning motion, W_n

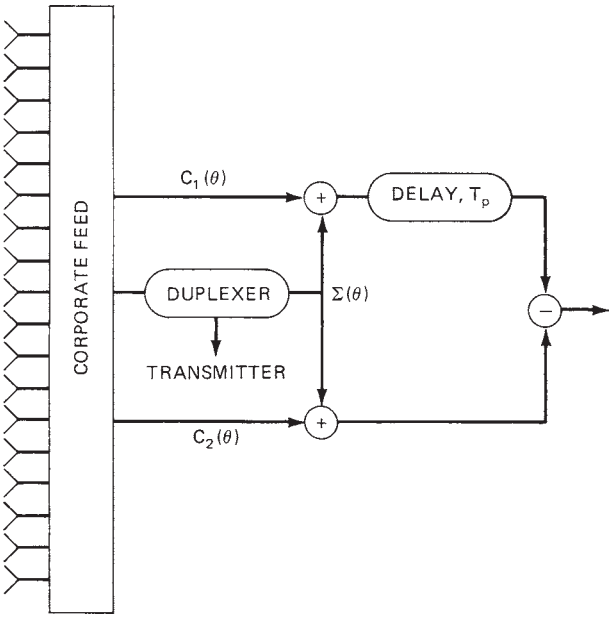


FIGURE 3.20 Optimized DPCA phase compensation

3.8 PLATFORM-MOTION COMPENSATION, FORWARD DIRECTION

The previous sections discussed the compensation for the component of platform motion parallel to the antenna aperture. TACCAR removes the average component of platform motion perpendicular to the aperture. The former Wheeler Laboratories developed the Coincident Phase Center Technique (CPCT)[15] to remove the spectral spread due to the velocity component perpendicular to the aperture and due to the component parallel to the aperture. Removal of the component parallel to the aperture uses the DPCA pattern synthesis technique described in Anderson,[8] which creates two similarly shaped illumination functions whose phase centers are physically displaced. Removal of the component perpendicular to the aperture is accomplished by a novel extension of this concept.

The first term of Eq. 3.3 for spectral width due to platform motion approaches zero as the antenna points ahead. However the second term of Eq. 3.3 dominates as the antenna approaches within a few beamwidths of the aircraft's ground track. In this region

$$f_d \approx \frac{4V_y}{\lambda} \sin 2\frac{\theta}{2} \approx \frac{V_y\theta^2}{\lambda} \tag{3.23}$$

which yields a single-sided spectrum that is significantly narrower than the spectrum abeam. For moderate platform speeds and lower-frequency (UHF) radars, this effect is negligible, and compensation is not required.

When it is necessary to compensate for this effect, the phase center of the antenna must be displaced ahead of the aperture and behind the aperture for alternate receive pulses so that the phase centers are coincident for a moving platform. This technique can be extended to more than two pulses by using the necessary phase-center displacements for each pulse. In order to maintain the effective PRF, the displacement must compensate for the two-way transmission path. To accomplish this displacement, near-field antenna principles are utilized. A desired aperture distribution function is specified. The near-field amplitude and phase are calculated at a given distance from the origin. If this field is used as the actual illumination function, a virtual aperture is created with the desired distribution function at the same distance behind the physical antenna. Figure 3.21*a*[15] shows the phase and amplitude distribution required to form a uniform virtual distribution displaced behind the physical aperture. It can be shown that if the phase of the illumination function is reversed $\phi' = -\phi$, the desired virtual distribution function is displaced ahead of the aperture, as shown in Figure 3.21*b*.

In practice, performance is limited by the ability to produce the required illumination function. As the displacement increases, a larger physical aperture size is required to produce the desired virtual aperture size owing to beam spreading. This can be seen in Figure 3.21. The effectiveness of the correction varies with elevation angle since the

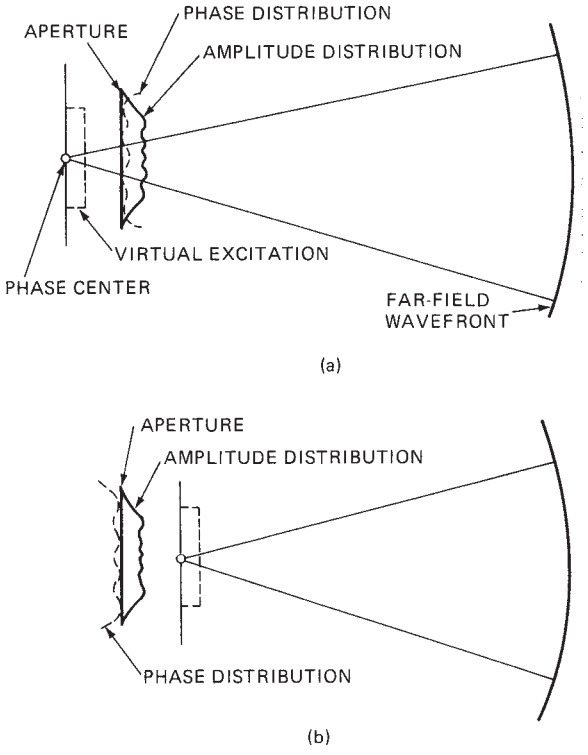


FIGURE 3.21 CPCT concept showing displacement of the phase center: (*a*) behind the physical aperture and (*b*) ahead of the physical aperture (*Courtesy of Hazeltine Inc.*[15])

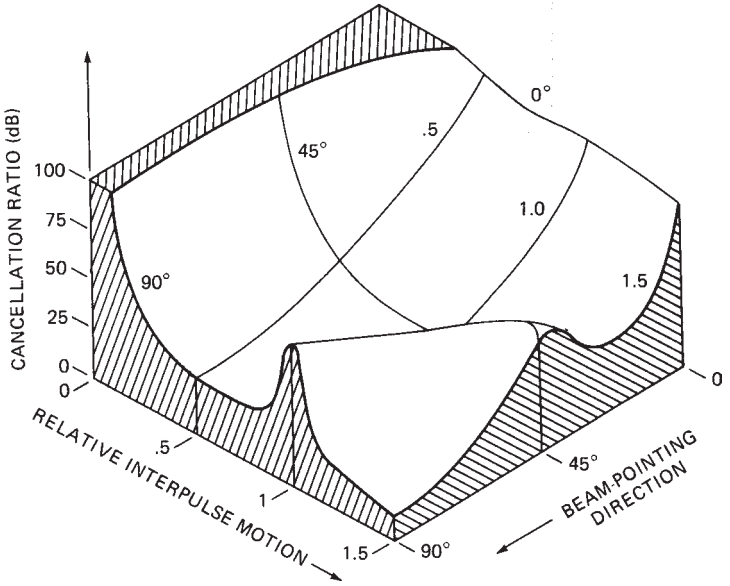


FIGURE 3.22 CPCT cancellation ratio, in decibels, as a function of relative interpulse motion and beam-pointing direction (*Courtesy of Hazeltine Inc.*[15])

actual displacement along the line-of-slight varies with elevation angle. This effect is more pronounced at higher aircraft speeds and higher radar frequencies. A change in the magnitude of the correction factor or even the compensation pattern with range, height, and velocity could be utilized to retain performance.

Figure 3.22 illustrates the theoretical MTI performance of a CPCT system as a function of beam-pointing direction and interpulse motion normalized to the interpulse motion used to design the compensation pattern. (*Cancellation ratio* is defined as the ratio of input clutter power to output clutter residue power.) The peak on the 90° axis is typical of the optimized DPCA performance illustrated in Figure 3.12.

3.9 SPACE-TIME ADAPTIVE MOTION COMPENSATION

Introduction. Several methods have been described to compensate for antenna motion. All these techniques are applied in the radar design phase for a specific set of operational parameters. Controls (usually automatic) are provided to adjust weights for operational conditions around the design value.

The development of digital radar technology and economical high-speed processors allows the use of dynamic space-time adaptive array processing (STAP),[16] whereby a set of antenna patterns that displace the phase center of the array both along and orthogonal to the array are continually synthesized to maximize the signal-to-clutter ratio. *Spatial adaptive array* processing combines an array of signals received at the same instant of time that are sampled at the different spatial locations corresponding

to the antenna elements. *Temporal adaptive array* processing combines an array of signals received at the same spatial location (e.g., the output of a reflector antenna) that are sampled at different instances of time, such as several interpulse periods for an adaptive MTI. Space-time adaptive array processing combines a two-dimensional array of signals sampled at different instances of time and at different spatial locations. STAP is a fairly broad topic that has applicability beyond this chapter on airborne MTI radar. The primary motivation for STAP is to improve clutter cancellation performance and to better integrate a radar's spatial processing (antenna sidelobe control and sidelobe jamming cancellation) with its temporal clutter cancellation processing.

The applicability of STAP to improving clutter cancellation must be assessed specifically in the context of the key performance limiters to airborne MTI radar clutter cancellation as described at the start of this chapter. STAP can improve a radar's motion compensation performance and is more robust than nonadaptive techniques in addressing generally non-dispersive errors in the radar front-end. STAP will not directly address clutter internal motion effects, antenna scanning motion effects, or other hardware stability impacts to clutter cancellation performance. Radar designers need to assess the key limitations in a specific application before jumping to the conclusion that STAP will improve performance.

STAP's ability to integrate clutter cancellation (temporal) and spatial interference cancellation can be quite important to many radar systems whether they typically have to deal with intentional jamming interference or unintentional (or casual) electromagnetic interference (EMI). STAP gets away from cascaded solutions such as analog sidelobe cancellers followed by digital DPCA and/or MTI filters—that do not generally create an optimum interference cancellation solution.

Optimal Adaptive Weights (McGuffin[17]). The optimal linear estimate is determined by requiring the adapted estimation error be orthogonal to the observed vector, $\underline{r}$. Steady-state conditions are assumed in this derivation, thus the condition for orthogonality is

$$E\{\underline{r}\ \varepsilon^*\} = 0 \tag{3.24}$$

where $E\{\ \}$ is the expectation, ε is the estimation error, and * is the complex conjugate. The adaptively weighted estimate is obtained by weighting the received signal vector by the estimate of the adaptive weights:

$$\hat{s} = \underline{\hat{w}}'\ \underline{r} \tag{3.25}$$

With d defined as the desired signal (a main-beam target), the estimation error is obtained from the following equation. Then, substituting Eq. 3.25 into 3.26 and solving for the adaptive weight estimate yields the desired condition for optimal adaptive weighting:

$$\varepsilon = \hat{s} - d = \underline{\hat{w}}'\ \underline{r} - d \tag{3.26}$$

$$E\{\underline{r}(d^* - \underline{r}'\ \hat{w})\} = 0 = E\{\underline{r}\ d^*\} - \underline{R}_r\ \hat{w}$$

or

$$\hat{w} = \underline{R}_r^{-1}\ E\{\underline{r}\ d^*\} \tag{3.27}$$

where $\underline{R}_r = E\{\underline{r}\ \underline{r}'\}$. The desired signal, d, can be expressed in terms of $\underline{s}$, the signal vector of a target located in the main beam, and $\underline{b}$, the unadapted beam weight vector: $d = \underline{b}'\ \underline{s}$. This is then substituted into Eq. 3.27:

$$\hat{w} = \underline{R}_r^{-1}\ \underline{R}_s\ \underline{b} \tag{3.28}$$

Equation 3.28 is equivalent to the minimum mean square error weight equation given by Widrow,[18] which has been shown[19,20] to be the optimum set that maximizes the signal-to-interference ratio. However, complex variables are employed here rather than real variables. The interference covariance matrix is further described in terms of the individual noise, jamming, clutter, and signal contributions:

$$\underline{R}_r = N\underline{I} + \underline{K}_Z + \underline{R}_S \tag{3.29}$$

where N is receiver noise power, $\underline{K}_Z$ is the covariance matrix for clutter (temporally correlated) plus jamming (spatially correlated), and $\underline{R}_s$ is the signal covariance matrix.

Taxonomy of STAP Architectures (Ward[21]). The application of the adaptive weight equation from Eq. 3.28 in a radar system provides numerous options and complications. The options range from a fully adaptive solution across all available antenna elements and all pulses in a coherent processing interval (CPI), to reduced degrees of freedom solutions in order to be practical. The fully adaptive solution also encounters problems in the real-world where the interference environment is not well behaved (e.g., homogenous clutter). In addition, Brennan's rule[22] indicates that to achieve an adaptive solution within 3dB of the optimum answer requires $2N$ (N is the number of degrees of freedom) independent interference samples contributing to the adaptive weight estimate. With antenna array sizes in tens to hundreds of elements and CPI lengths of tens to hundreds of pulses, the number of degrees of freedom can quickly get quite large, resulting in not only fairly complex adaptive weight processing but also the more difficult problem of obtaining adequate sample support from clutter and jamming interference for a given adaptive weight solution.

As such, it is important to explore various STAP architecture options imbedded in a radar design solution. To begin, a fully adaptive array architecture is shown in Figure 3.23. This is for a linear array antenna with a distributed transmitter and digital receivers connected to each antenna element. The adaptive weight solution is developed based on at least $2 \times N \times M$ vector samples ($\underline{r}$) of length M (antenna elements) by N (pulses). The adaptive weight solution is developed and applied to the received signals from the same antenna elements and pulses of data. The adaptive weighted response is typically processed through doppler filtering (coherent integration) prior to detection processing.

Ward[21] describes the possible STAP architectures in the context of a generalized transformation matrix followed by the associated STAP processing. The four categories of STAP architectures are organized in Figure 3.24. The trades for an appropriate STAP design solution must be made in the context of the type and size of the antenna aperture under consideration, the waveforms under consideration—particularly the number of pulses per CPI—and most importantly, the interference to be cancelled (clutter and jamming). In general, for the transformation and degrees of freedom reduction to be useful, the resultant degrees of freedom must be greater than the interference rank.

Pre-Doppler, Elemental Antenna STAP. Conceptually, the simplest reduction in degrees of freedom is obtained by reducing the number of temporal degrees of

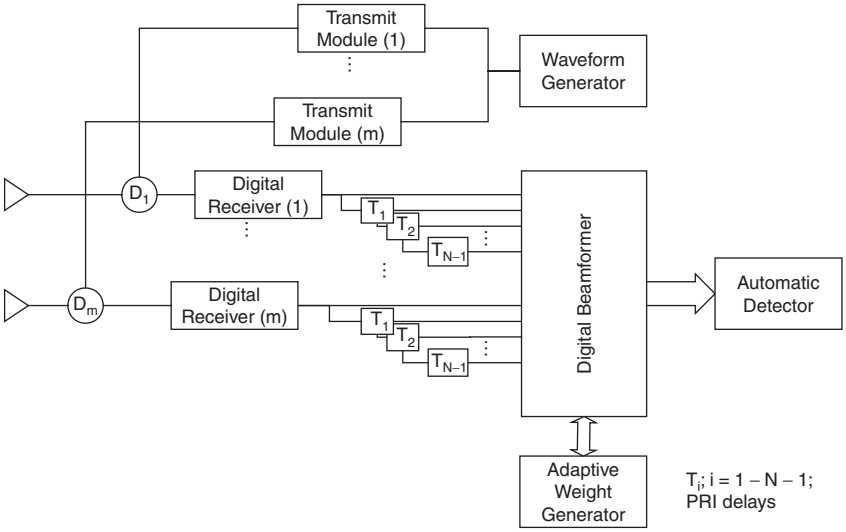


FIGURE 3.23 STAP radar block diagram

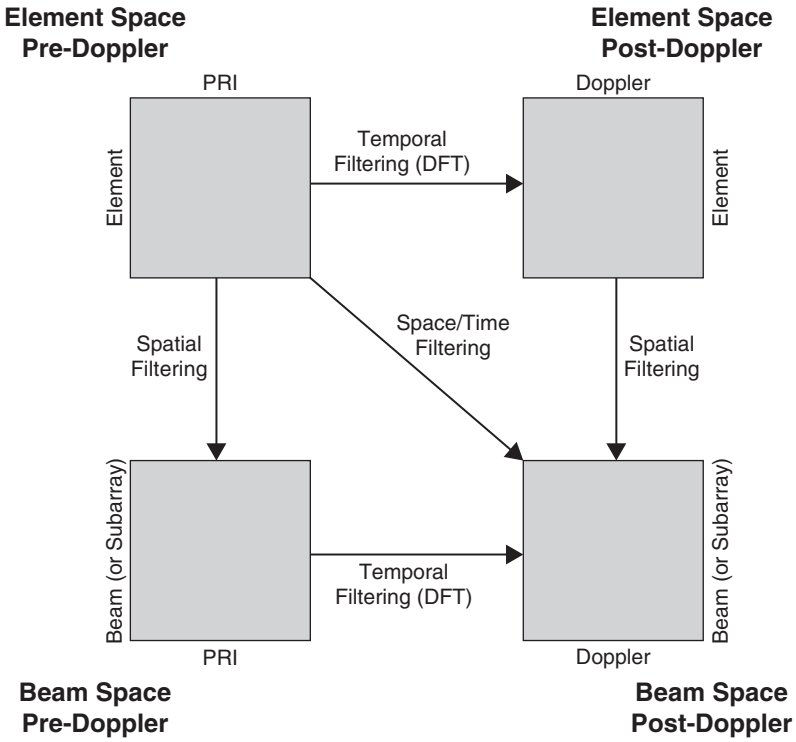


FIGURE 3.24 Reduced dimension STAP architectures

freedom in STAP while still processing the full aperture spatially. This is similar to a conventional MTI (or DPCA) architecture cascaded with doppler filtering. We call this architecture a pre-doppler, elemental-level STAP architecture. For a three-pulse version of this architecture, there are 3M degrees of freedom. In this architecture, platform motion compensation takes the general form of adjusting the antenna's phase center over the three temporally separated beams.

A basic block diagram of a radar incorporating pre-doppler, elemental-level space-time adaptive array processing is shown in Figure 3.25. An individual duplexer is placed between each transmitter's channelized output and its corresponding antenna element. Provision could be included for electronic beam steering using high-power phase shifters or transmit modules with low-power beam steering.

On receive, each duplexer output is sent to its own digital receiver. The digital receiver outputs are passed through PRI delays to yield temporally displaced data samples. A full complement of elements and time-delayed signals are sampled and used to generate the adaptive weights. Various algorithms are possible to generate the estimate of the adaptive weights from Eq. 3.28. The fairly simple Least Mean Squared algorithm generally yields fairly slow convergence rates. Other algorithms[19,23] can speed up the adaptation rate, but a more complex mechanization is required. Examples include a Recursive Least Squared algorithm, Q-R decomposition with Gram-Schmidt orthogonalization, or a Householder Transformation. The adaptive weights are then applied to the received signals and beamformed to generate three sum channel detection beams: undelayed, one-PRI delayed, and two-PRI delayed beams. These beams are, in turn, added together to form the final STAP weighted detection beam.

A simplistic view of how these three beams perform motion compensation is illustrated in Figure 3.26 for the case where the aperture is parallel with the radar's platform velocity vector. The first pulse return's phase center is advanced by aperture weighting, the second pulse return's phase center is essentially unchanged from the quiescent weights, and the third pulse return's phase center is retarded by aperture weighting. Given ideal antenna patterns, and an aperture large enough to adjust the phase centers

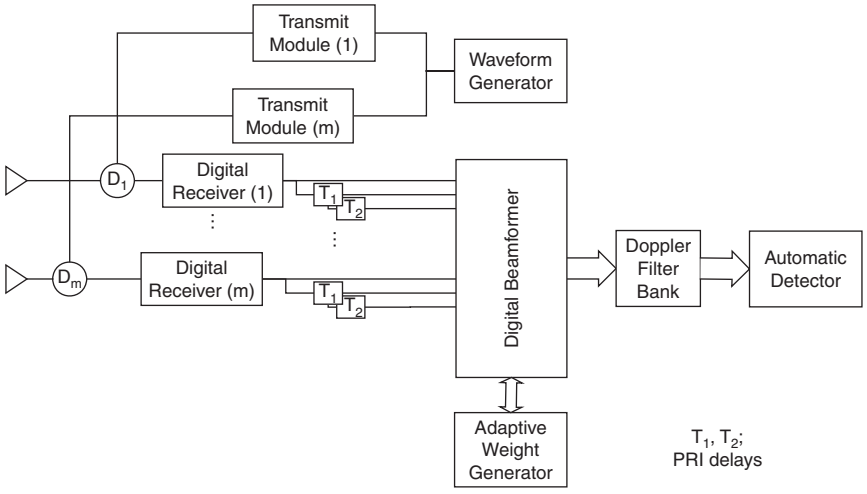


FIGURE 3.25 STAP block diagram: element space pre-doppler element space architecture

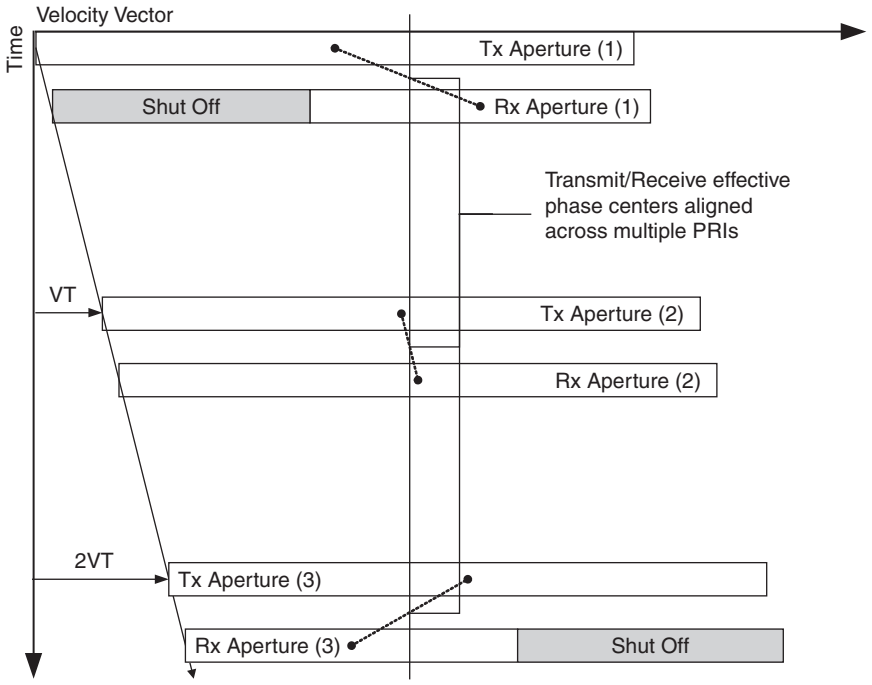


FIGURE 3.26 Aperture control for platform motion compensation

for the given platform motion, these three apertures appear as if they are stationary with respect to each other. Clutter cancellation across these three pulses is no longer limited by platform motion effects—the primary goal of platform motion compensation techniques.

Of course, this simplest condition is only illustrative, as generally the antenna elements do not behave exactly the same, and the platform motion compensation must deal with motion not only in the plane of the aperture but also orthogonal to the aperture.

Pre-Doppler, Beam-Space STAP. The first type of transformation to be considered is spatially oriented, resulting in beam-space STAP architectures. This transformation is typically required for many large apertures. The transformations can range from simple column beamforming to overlapped subarrays to beam-space transformations such as a Butler matrix. The general goal is to reduce the spatial degrees of freedom, while still providing access to array responses that allow for adequate clutter cancellation and beams that can be used to cancel directional interference as well. The resulting beam responses must span the clutter and jamming interference spatially in order for this type of transformation to be effective. For example, if a radar's clutter cancellation performance is driven by main-beam clutter residue due to platform motion effects, the beam responses must span the radar's main-beam and provide degrees of freedom to allow for motion compensation in the array main-beam. In addition, to cancel direction interference (jamming or casual EMI), the beam responses

must also span the spatial directions of that interference. An example of a simple transformation of this type would be sidelobe canceler architecture where the beam transformation would generate a sum channel main beam and select elements from the aperture as sidelobe cancellers.

Post-Doppler, Element Antenna STAP. The second type of transformation leads to what are called post-doppler STAP architectures. As the name implies, the antenna element signals are first doppler filtered and then processed through STAP. The motivation for this type of architecture is that the resultant STAP solutions can independently address a subset of the clutter interference problem isolated to clutter that remains in a single doppler filter. This technique may be more effective for radar systems where the clutter environment and waveform selection lead to unambiguous clutter returns within the radar's PRF. Two example conditions, the first with ambiguous doppler clutter and the second with unambiguous doppler clutter, are shown in Figure 3.27. The figure shows those antenna angles where the clutter doppler response remains after filtering through a single doppler filter. Figure 3.27*a* shows the response for an ambiguous PRF of 300 Hz, and Figure 3.27*b* shows the response for an unambiguous PRF of 2000 Hz for a UHF radar. This figure highlights that even with doppler processing, a given doppler filter may still include clutter returns from a number of discontiguous angular intervals. The advantages of this transformation from PRI to doppler space on overall STAP performance versus a pre-doppler architecture are more dramatic in the unambiguous doppler clutter case.

PRI-staggered doppler filter outputs are required to maintain a set of temporal degrees of freedom in this architecture. The block diagram is modified to that shown in Figure 3.28, with multiple doppler filter banks on each antenna element and PRI delay.

Post-Doppler, Beam Space STAP. The final category results from implementing both doppler and spatial transformations prior to STAP processing.

The appropriate architecture solution depends upon the radar design constraints. The number of antenna elements and beamforming requirements are key drivers in the

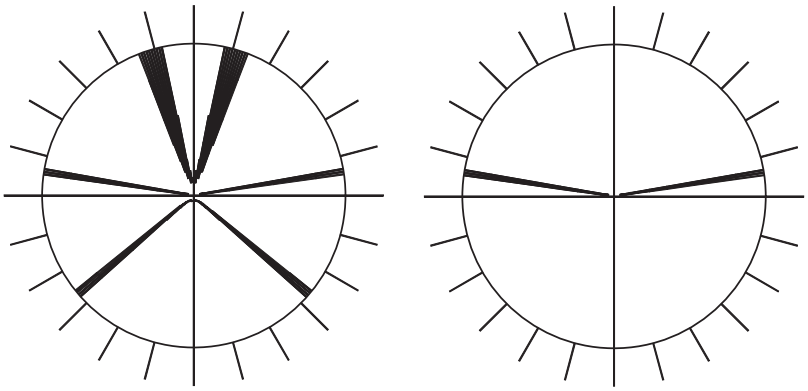


FIGURE 3.27 Antenna pointing angles where clutter doppler map to a single doppler filter's passband

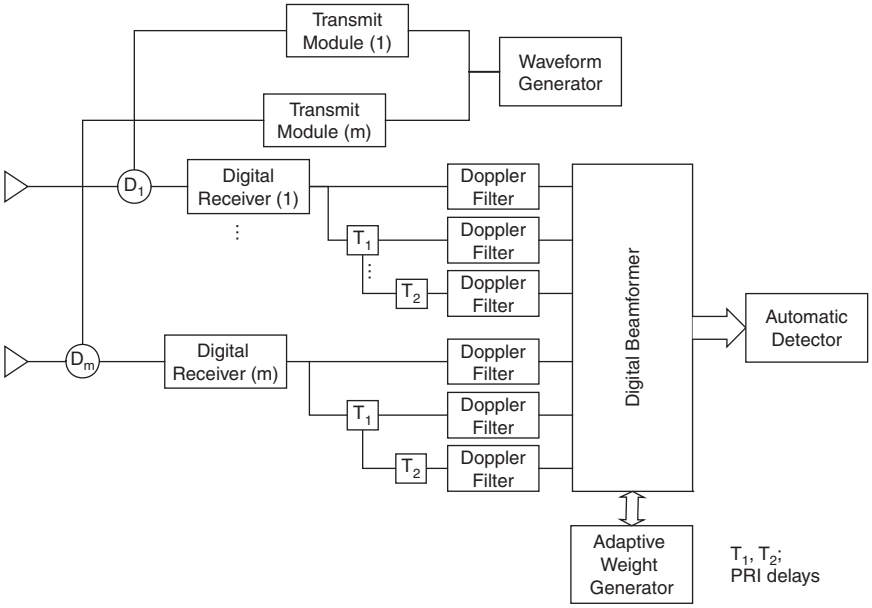


FIGURE 3.28 Element space post-doppler STAP architecture

decision whether to transform from elements to beams or subarrays. The waveforms and clutter cancellation requirements are key drivers in the decision whether to perform STAP on signals before or after doppler filtering. In addition, the overall transformation decisions to reduce degrees of freedom are driven by the interference rank for the radar problem. One caution in the design process is that if the transformation is fixed in the radar design, it is important to have excess degrees of freedom beyond the total interference rank.

Implementation Considerations. As discussed above, transformations and techniques to reduce the number of degrees of freedom in the STAP solution are important not only due to processing requirements but also because of the need for sample support on the order of two times the number of degrees of freedom for adequate STAP performance.

The basic hardware requirements for good clutter cancellation remain unchanged from conventional clutter cancellation architectures—low phase noise, low pulse jitter, etc. The requirements on the hardware may become more stringent because the STAP architecture allows the radar designer to achieve higher theoretical clutter cancellation performance levels. In addition to the above temporally based hardware requirements, there are also second-order spatially based hardware requirements. As illustrated in Figure 3.26, platform motion compensation results in different aperture weighting for successive pulses in a STAP solution. Although generally speaking, well-matched spatial channels (antenna and receiver) are driven by jamming cancellation and antenna sidelobe levels, a second-order requirement results from the need for

platform motion compensation. If antenna and receiver channels are not well matched, the resultant sum channel beams formed from different aperture illumination functions (Figure 3.26) will not be matched well enough to provide main-beam and sidelobe clutter cancellation.

Performance Comparisons. Given the number of STAP architectures and corresponding radar system design solutions, general STAP performance comparisons are difficult to come by. In general, STAP provides a robust solution to deal with clutter and jamming interference and helps alleviate hardware mismatch effects within reason (amplitude and phase adjustments are applied to antenna element and time displaced returns). Generally to address time-delay adaptive weighting, more complexity is required with a third dimension for adaptive weights—"fast-time" or returns from adjacent sampled range cells. This extension can be extremely computationally intensive and further burden the sample support problem alluded to previously.

When evaluating a radar design and trading off various waveforms and STAP processing techniques, it is important to include in the analysis key drivers such as signal bandwidth, clutter internal motion, platform motion, antenna scanning motion, the amount of sample support available from nonhomogenous and nonstationary clutter environments, and other effects such as large target samples effecting the adaptive weight solution.

3.10 EFFECT OF MULTIPLE SPECTRA

An airborne search-radar system may be operated at an altitude so that the radar horizon is approximately at the maximum range of interest. This results in sea or ground clutter being present at all ranges of interest. Other clutter sources such as rain and chaff may coexist with the surface clutter. In most instances, these sources are moving at a speed determined by the mean wind aloft and have a mean doppler frequency significantly different from that of the surface clutter. If the MTI filter is tracking the surface clutter, the spectra of the sources with a different mean doppler frequency lie in the passband of the MTI filter. A 20-kt differential in a UHF system corresponds to 30 Hz, which would generally be outside of the traditional AMTI notch filter in a 300 Hz PRF system. A single-delay secondary canceler can be cascaded with either a single-delay or a double-delay primary canceler. The primary canceler tracks the mean surface velocity and rejects surface clutter. The single-delay canceler tracks the secondary source and rejects it. Since the pass and rejection bands of the two cancelers overlap, the MTI improvement factor for each clutter source is a function of their spectral separation.

Figure 3.29 shows the improvement factor for a double canceler, which consists of two single cancelers, each tracking one of the spectra. It can be seen that as the separation varies from 0 to 1/2 of the PRF, the performance degrades from that equivalent to a double canceler to the performance of a single canceler at half of the PRF.

The triple canceler has a double-delay canceler tracking the primary spectra and a single-delay canceler tracking the secondary spectra. The performance of the primary system varies from that of a triple canceler to a level less than that of a double canceler. The secondary-system performance varies from that of a triple canceler to a performance level lower than that of a single canceler.

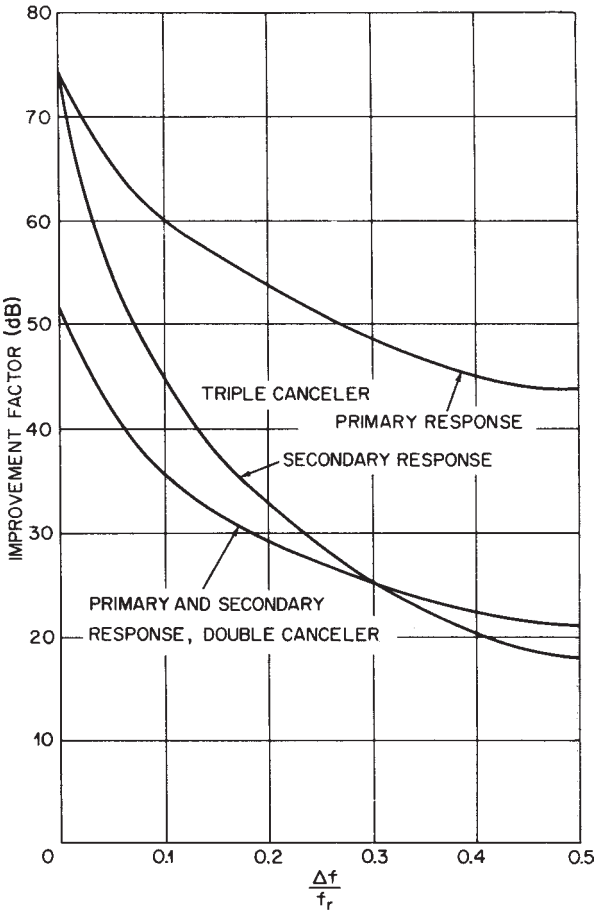


FIGURE 3.29 MTI improvement factor for a double-notch canceler tracking two spectra as a function of the normalized spectra separation $\Delta f/f_r$. Normalized spectral width $\sigma_c/f_r = 0.01$.

3.11 EXAMPLE AMTI RADAR SYSTEM

The AN/APY-9 radar, developed by Lockheed Martin for the U.S. Navy, is an example of an AMTI radar system utilized for an airborne early warning radar mission. Key features of this system include a solid-state distributed transmitter, a mechanically and electronically scanned rotating antenna, digital receivers, space-time adaptive processing, digital pulse compression, and coherent integration and auxiliary processing aimed at supporting the STAP sample selection process.

The AN/APY-9 radar addresses the AEW radar surveillance coverage requirements discussed at the beginning of this chapter, utilizing a mechanically and electronically steerable antenna located in a rotodome. There are three scanning modes of operation:

(1) mechanically scanned with an operator-selectable scan rate, (2) azimuth electronically scanned with the mechanical boresite provided as an input to the radar, and (3) mechanically scanned with additional electronic scanning within an operator-selectable azimuth region.

The transmit waveform includes TACCAR modulation to center mainbeam clutter at zero doppler frequency. However, because the radar implements adaptive clutter cancellation (STAP), the requirements on TACCAR are significantly less complex than for legacy radar systems. There is no need to include closed loop adjustments to the TACCAR modulation frequency. The optimization of the AMTI clutter cancellation filter is achieved in the STAP processing as opposed to adjusting the location of main-beam clutter to fit a fixed AMTI filter.

In order to implement STAP and electronic scanning in this radar, all 18 elements of the phased array antenna are processed on transmit and receive. The solid-state transmitter provides low-power phase shift control for electronic steering followed by power amplification in each of 18 channels. These are connected to the 18 elements of the phased array through an 18-channel rotary coupler. The transmit/receive isolation on all 18 channels is provided through circulators. The 18 channels are processed separately through 18 receivers, finally feeding the STAP subsystem with 18-digital baseband signals.

The radar performs platform motion compensation electronically as part of the STAP architecture. The radar implements an element-space pre-doppler STAP architecture. Adaptive weights are generated and applied to the 18 receive channels, forming three beams (Sum, Delta_{az}, and Omni) by weighting and summing the 18 receive channels over three pulses to provide simultaneous clutter and jamming cancellation. The adaptive weight algorithm is matched to the radar's operating parameters and is augmented with adaptive knowledge–aided sampling schemes to maximize performance in a complex, heterogeneous clutter and jamming interference environment. Doppler filtering is performed after digital beamforming.

Other functions discussed in this chapter are not required for this radar application because they do not limit performance. Examples include scanning motion compensation and multiple spectra AMTI clutter cancellation.

REFERENCES

1. R. C. Emerson, "Some pulsed doppler MTI and AMTI techniques," Rand Corporation Rept. R-274, DDC Doc. AD 65881, March 1, 1954. (Reprinted in Reference 6.)
2. T. S. George, "Fluctuations of ground clutter return in airborne radar equipment," *Proc. IEE* (London), vol. 99, pt. IV, pp. 92–99, April 1952.
3. F. R. Dickey, Jr., "Theoretical performance of airborne moving target indicators," *IRE Trans*., vol. PGAE-8, pp. 12–23, June 1953.
4. R. S. Berkowitz (ed.), *Modern Radar: Analysis, Evaluation and System Design*, New York: John Wiley & Sons, 1966.
5. D. K. Barton, *Radar Systems Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1964.
6. D. C. Schlerer (ed.), *MTI Radar*, Norwood, MA: Artech House, Inc., 1978.
7. F. R. Dickey, Jr. and M. M. Santa "Final report on anticlutter techniques," General Electric Company Rept. R65EMH37, March 1, 1953.
8. D. B. Anderson, "A microwave technique to reduce platform motion and scanning noise in airborne moving target radar," *IRE WESCON Conv. Rec*., vol. 2, pt. 1, 1958, pp. 202–211.

9. "Final engineering report on displaced phase center antenna," vol. 1, March 26, 1956; vols. 2 and 3, April 18, 1957, General Electric Company, Schenectady, NY.
10. H. Urkowitz, "The effect of antenna patterns on performance of dual antenna radar moving target indicators," *IEEE Trans*., vol. ANE-11, pp. 218–223, December 1964.
11. G. N. Tsandoulis, "Tolerance control in an array antenna," *Microwave J*., pp. 24–35, October 1977.
12. K. G. Shroeder, "Beam patterns for phase monopulse arrays," *Microwaves*, pp. 18–27, March 1963.
13. R. S. Grissetti, M. M. Santa, and G. M. Kirkpatrick, "Effect of internal fluctuations and scanning on clutter attenuation in MTI Radar," *IRE Trans*., vol. ANE-2, pp. 37–41, March 1955.
14. G. A. Andrews, "Airborne radar motion compensation techniques: Optimum array correction patterns," Naval Res. Lab. Rept. 7977, March 16, 1976.
15. A. R. Lopez and W. W. Ganz, "CPCT antennas for AMTI radar, vol. 2: Theoretical study," Air Force Avionics Lab. Rept. WL1630.22, AD 51858, June 1970. (Not readily available.)
16. L. E. Brennan., J. D. Mallett, and I. S. Reed, "Adaptive arrays in airborne MTI radar," *IEEE Trans*., vol. AP-24, pp. 607–615, September 1976.
17. A. L. McGuffin, "A brief assessment of adaptive antennas with emphasis on airborne radar," General Electric Company, Aircraft Equipment Division, August 1981.
18. B. Widrow and S. D. Stearns. *Adaptive Signal Processing,* New Jersey: Prentice-Hall, Inc., 1985.
19. S. P. Applebaum, "Adaptive arrays," *IEEE Trans*., vol. AP-24, pp. 585–598, September 1976.
20. L. E. Brennan, E. L. Pugh, and I. S. Reed, "Control loop noise in adaptive array antennas," *IEEE Trans*., vol. AES-7, March 1971.
21. J. Ward, "Space time adaptive processing for airborne radar," MIT Lincoln Laboratory Technical Report #1015, December. 13, 1994.
22. L. E. Brennan and F. M. Staudaher, "Subclutter visibility demonstration," Technical Report RL-TR-92-21, Adaptive Sensors Incorporated, March 1992.
23. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, New York: John Wiley & Sons, 1980.

Chapter 4
Pulse Doppler Radar*

John P. Stralka
William G. Fedarko

Northrop Grumman Corporation

4.1 CHARACTERISTICS AND APPLICATIONS

The primary benefit of pulse doppler radar is its ability to detect small-amplitude moving target returns against an overwhelmingly large-amplitude clutter background.

Nomenclature. Radars that rely on the doppler effect to enhance target detection are called *doppler radars*.[1] The doppler effect manifests itself when there is a relative range rate, or radial velocity, between the radar and the target. When the radar's transmit signal is reflected from such a target, the carrier frequency of the return signal will be shifted. Assuming a monostatic radar (i.e., collocated transmitter and receiver), the roundtrip distance is twice the distance between the transmitter and the target. The doppler frequency shift f_d is a function of the carrier wavelength λ and the relative radial velocity (range rate) between the radar and the target V_{relative}, and is written as $f_d = -2V_{\text{relative}}/\lambda$, where $\lambda = c/f$ is the wavelength, c is the speed of light, and f is the carrier frequency. When the target is moving away from the radar, the relative radial velocity, or range rate, is defined to be positive and results in a negative doppler shift.

Doppler radars can be either continuous wave (CW)† or pulsed radars. CW radars simply observe the doppler shift between the carrier frequency of the return signal relative to the transmit signal. Pulsed systems measure doppler by using a coherent train of pulses where there is a fixed or deterministic phase relationship of the carrier frequency between each successive radio frequency (RF) pulse. Coherence concentrates the energy in the frequency spectrum of the pulse train around distinct spectral lines, separated by the pulse repetition frequency (PRF). This separation into spectral lines allows for discrimination of doppler shifts.

Doppler radars using pulsed transmissions are more complex than CW radars, but they offer significant advantages. Most important is the time gating of the receiver.

* David H. Mooney and William A. Skillman wrote this chapter for the first edition (1970). William H. Long joined the authors for the second edition (1990). John P. Stralka and William G. Fedarko updated the material for this edition.

† To assist the reader, abbreviations used throughout this chapter are defined in a list at the end of the chapter.

Time gating allows the blanking of direct transmitter leakage into the receiver. This permits the use of a single antenna for transmit and receive, which otherwise would not be feasible for CW radar due to excessive transmit/receive isolation requirements. Pulsed radars can also use range gating, a specific form of time gating, which divides the interpulse period into cells or *range gates*. The duration of each cell is typically less than or equal to the inverse of the transmit pulse bandwidth. Range gating helps eliminate excess receiver noise from competing with target returns and allows range measurement with pulse delay ranging (i.e., measuring the time between transmission of a pulse and reception of the target echo).

Pulsed transmission doppler radars have historically been categorized as *moving target indication (MTI)* or *pulse doppler*. MTI typically eliminates clutter by passing the received returns from multiple coherent pulses through a filter with a stopband placed in spectral regions of heavy clutter concentrations. Moving targets with doppler frequencies outside the stopband are passed onto detection processing. Pulse doppler radars, on the other hand, resolve and enhance targets within a particular doppler band while rejecting clutter and other returns outside the doppler band of interest. This is typically accomplished with a contiguous bank of doppler filters formed between two of the coherent pulse train's spectral lines, one of which is the central line. Range gating precedes the doppler filter bank. The bandwidth of each doppler filter is inversely proportional to the duration of the coherent pulse train that is processed to form the doppler filter bank. This process forms a matched filter to the entire pulse train.[2,3]

MTI and pulse doppler radars share the following characteristics:

- Coherent transmission and reception; that is, each transmitted pulse and the receiver local oscillator are synchronized to a free-running, highly stable oscillator.
- Coherent processing to reject main-beam clutter, enhance target detection, and aid in target discrimination or classification.

MTI radars can also be implemented using a doppler filter bank, blurring the historic delineation between MTI and pulse doppler radars. As a result, this book will define MTI radars as those radars whose PRF is sufficiently low enough to provide an unambiguous range measurement, via pulse delay ranging, over the radar's instrumented range. The unambiguous range R_u is given by $c/(2f_R)$, where c is the speed of light and f_R is the PRF. Radars with PRFs that result in range ambiguities within the range coverage of interest will be referred to as pulse doppler radars and will be the focus of this chapter.

Applications. Pulse doppler is applied principally to radar systems requiring the detection of moving targets in a severe clutter environment. Table 4.1 lists typical applications and requirements.[4–12] This chapter will deal principally with airborne applications, although the basic principles can also be applied to the surface-based case. Only monostatic radars will be considered.

PRFs. Pulsed radars that employ doppler are divided into three broad PRF categories: low, medium, and high. A low-PRF radar is one in which the ranges of interest are unambiguous while the radial velocities (doppler frequencies) are usually highly ambiguous. As discussed previously, this type of radar is called *moving target indication (MTI)*. MTI radars are generally not categorized as pulse doppler radars, although the principles of operation are similar.[13]

TABLE 4.1 Pulse-Doppler Applications and Requirements

Radar Application	Requirements
Airborne or spaceborne surveillance	Long detection range; accurate range data
Airborne interceptor or fire control	Medium detection range; accurate range, velocity, and angle data
Ground-based surveillance	Medium detection range; accurate range data
Battlefield surveillance (slow-moving target detection)	Medium detection range; accurate range, velocity data
Missile seeker	Short detection range; accurate velocity and angle rate data; may not need true range information
Surface-based weapon control	Short range; accurate range, velocity data
Meteorological	Good velocity resolution
Missile warning	Short detection range; very low false-alarm rate

The converse of a low-PRF radar is a high-PRF radar that can measure doppler unambiguously over the span of radial velocities of interest, but is usually highly ambiguous in range. A medium-PRF radar has both range and doppler ambiguities.[14–17] A blend of medium and high PRF, known as high-medium PRF (which will be discussed later), is characterized as having only a single-ambiguity for the radial velocities of interest. For this chapter, a pulse doppler radar is characterized as having a PRF anywhere within the medium to high PRF regime that results in ambiguous range measurements during a coherent processing interval.

A comparison of MTI and pulse doppler radars is shown in Table 4.2. Previously undefined terms will be defined throughout the chapter. The table assumes an airborne radar application designed to detect other aircraft. Such an application is commonly referred to as *air-to-air*.

TABLE 4.2 Comparison of MTI and Pulse Doppler Radars for Air-to-Air

	Advantages	Disadvantages
Low PRF *MTI* range unambiguous doppler ambiguous	Can sort clutter from targets on basis of range. Front-end sensitivity time control (STC) suppresses sidelobe detections at short ranges and reduces dynamic range requirements.	Multiple blind speeds. Usually does not measure radial target velocity. Poor ground-moving target rejection.
Medium PRF *Pulse Doppler* range ambiguous doppler ambiguous	Performance at all target aspects. Good ground-moving target rejection. Measures radial velocity. Less range eclipsing than in high-PRF.	Sidelobe clutter can limit performance. Ambiguity resolution required. Low antenna sidelobes necessary. Rejection of sidelobe returns of discrete ground targets needed.
High PRF *Pulse Doppler* range ambiguous doppler unambiguous	Allows thermal noise-limited detection of targets with high radial velocities. Single doppler blind zone at zero velocity. Good ground-moving target rejection. Measures radial velocity.	Limited low radial velocity target detection. Range eclipsing. Large number of range ambiguities preclude pulse delay ranging. High stability requirements due to range folding.

TABLE 4.3 Typical Values for an X-band (10 GHz) Airborne Fire-Control Radar

Pulse Doppler Waveform	PRF	Transmit Duty Cycle
Medium PRF	10–40 kHz	5–10%
High-medium PRF	60–100 kHz	10–20%
High PRF	120–300 kHz	15–50%

Table 4.3 provides the span of PRFs and corresponding transmit *duty cycles* (ratio of transmit pulse width to interpulse period) for the various pulse doppler waveforms used in a X-band airborne fire-control radar. Keep in mind that the operating frequency of the radar, along with its required range and radial velocity coverage, determines whether a PRF is considered medium, high-medium, or high. Also, modern multifunction radars are typically capable of utilizing waveforms from the various PRF categories in order to carry out their diverse missions.

Pulse Doppler Spectrum. The transmitted spectrum of a pulse doppler radar consists of discrete lines at the carrier frequency f_0 and at sideband frequencies $f_0 \pm if_R$, where f_R is the PRF and i is an integer. The envelope of the spectrum is determined by the pulse shape. For the rectangular pulses usually employed, a $\sin(x)/x$ spectrum is obtained.

Using a constant-velocity airborne radar, the received spectrum from a stationary target has lines that are doppler-shifted proportionally to the radial velocity between the radar platform and the target. The two-way doppler shift is given by $f_d = (2V_R/\lambda)\cos(\psi_0)$, where λ is the radar wavelength, V_R is the radar platform speed, and ψ_0 is the angle between the velocity vector and the line of sight to the target. (Note that the relative radial velocity (range rate) to the stationary target is $V_{\text{relative}} = -V_R\cos(\psi_0)$, which makes the later equation for doppler shift consistent with the one presented at the beginning of the chapter.) Illustrated in Figure 4.1 is the received pulsed spectrum with returns from distributed clutter, such as the ground or weather, and from discrete targets, such as aircraft, automobiles, tanks, etc.

Figure 4.2 shows the unfolded spectrum (i.e., no spectral foldover from adjacent PRF lines) in the case of horizontal motion of the radar platform, with a speed V_R. The clutter-free region is defined as that portion of the spectrum in which no ground clutter can exist. (A clutter-free region usually does not exist with medium PRFs due to doppler folding.) The sidelobe clutter region, $4V_R/\lambda$ in width, contains ground clutter power from the sidelobes of the antenna, although this clutter power may be below the noise level in part of the region. The main-beam clutter region, located at $f_0 + (2V_R/\lambda)\cos(\psi_0)$, contains the strong return from the main beam of the antenna

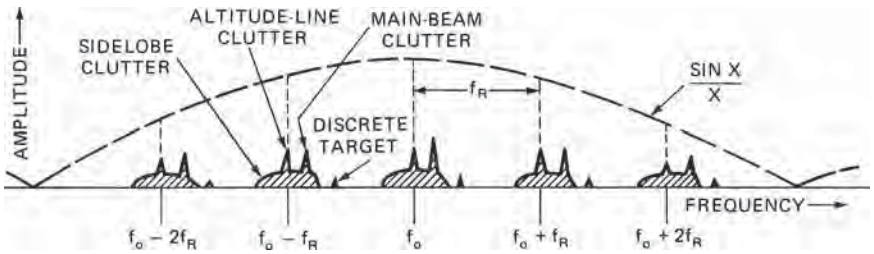


FIGURE 4.1 Clutter and target frequency spectrum from a horizontally moving platform

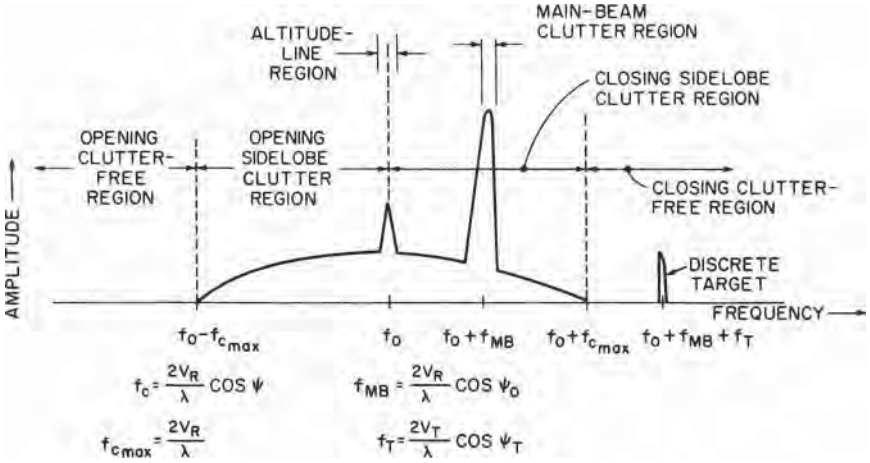


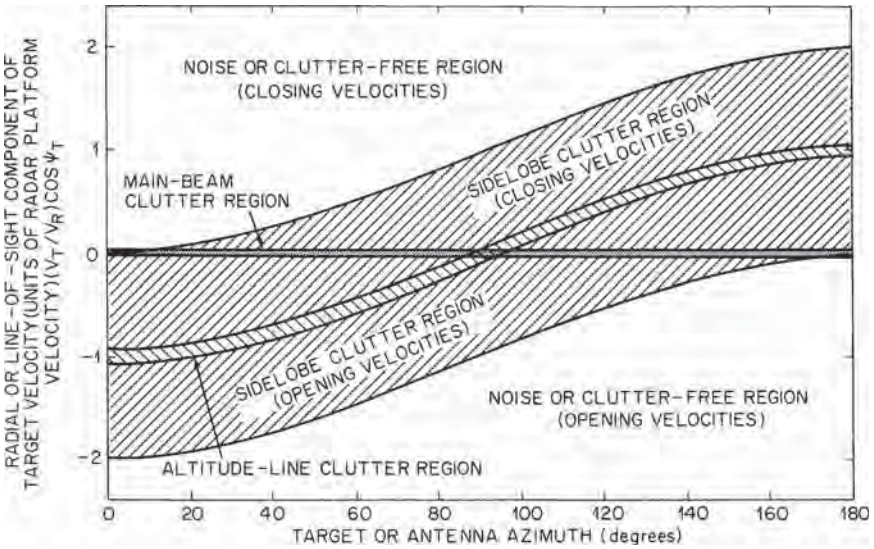
FIGURE 4.2 Unfolded spectrum (with no clutter positioning)

striking the ground at a scan angle of ψ_0, measured from the velocity vector. Rain and chaff clutter may also be large when the main beam illuminates a rain or chaff cloud. Motion due to winds may displace and/or spread the return in frequency.

Altitude-line clutter is due to the radar return from ground clutter at near normal incidence directly below the radar platform, and is at zero doppler if there is no vertical component of platform velocity. A discrete target return in the main beam is shown at $f_T = f_0 + (2V_R/\lambda)\cos(\psi_0) + (2V_T/\lambda)\cos(\psi_T)$, where the target speed is V_T, with an angle ψ_T between the target velocity vector and the radar target line of sight. The components of the spectrum shown in Figure 4.2 will also vary with range, as discussed later. (Note that the direction of $V_T\cos(\psi_T)$ is assumed to be the opposite of $V_R\cos(\psi_0)$ resulting in a relative range rate of $V_{\text{relative}} = -V_T\cos(\psi_T) - V_R\cos(\psi_0)$, which is consistent with the definition for doppler shift stated at the beginning of the chapter.)

Figure 4.3 illustrates the various clutter doppler frequency regions as a function of the antenna main-beam azimuth and relative radar and target velocities, again for an unfolded spectrum. The ordinate is the radial or line-of-sight component of target velocity in units of radar platform velocity, so the main-beam clutter region is at zero velocity and the sidelobe clutter region frequency boundaries vary sinusoidally with antenna azimuth. Thus, the figure shows the doppler regions in which the target becomes clear of sidelobe clutter. For example, if the antenna main-beam azimuth angle is at zero, any head-on target ($V_T\cos(\psi_T) > 0$) is clear of sidelobe clutter, whereas if the radar is in trail behind the target ($\psi_T = 180°$ and $\psi_0 = 0°$), the target's radial velocity has to be greater than twice that of the radar to become clear of sidelobe clutter.

The sidelobe clear and clutter regions can also be expressed in terms of the aspect angle with respect to the target, as shown in Figure 4.4.[18] Here, collision geometry is assumed in which the radar and target aircraft fly straight-line paths toward an intercept point; the look angle of the radar ψ_0 and the aspect angle of the target ψ_T are constant for a given set of radar and target speeds V_R and V_T, respectively. The center of the diagram is the target, and the angle to the radar on the circumference is the aspect angle. The aspect angle and look angles satisfy the equation $V_R \sin(\psi_0) = V_T \sin(\psi_T)$,



NOTE: Width of altitude-line and main-beam clutter regions varies with conditions; azimuth is measured from radar platform velocity vector to the antenna boresight or to the line of sight to the target; horizontal-motion case.

FIGURE 4.3 Clutter and clutter-free regions as a function of target velocity and azimuth

which is defined as a collision course. The target aspect angle is zero for a head-on condition and 180° for a tail chase. The aspect angle corresponding to the boundary between the sidelobe clutter region and the sidelobe clear region is a function of the relative radar-target velocity ratio and is shown in Figure 4.4 for four cases. Case 1 is where the radar and target speeds are equal and the target can be seen clear of sidelobe clutter in a head-on aspect out to 60° on either side of the target's velocity vector. Similarly, Cases 2 through 4 show conditions where the target's speed is 0.8, 0.6, and 0.4 times the radar's speed, in which case the target can be seen clear of sidelobe clutter over a region of up to ±78.5° relative to the target's velocity vector. Again, these conditions are for an assumed collision course. As is evident, the aspect angle of the target clear of sidelobe clutter is always forward of the beam aspect.

Ambiguities and PRF Selection. Pulse doppler radars are ambiguous in range and possibly doppler. As mentioned earlier, the unambiguous range R_u is given by $c/(2f_R)$, where c is the speed of light and f_R is the PRF.

If the airborne target radial velocity to be observed is between $V_{T,\text{max,opening}}$ for opening targets (positive range rate) and $-V_{T,\text{max,closing}}$ for closing targets (negative range rate), then the minimum value of PRF, $f_{R\,\text{min}}$, which is unambiguous in velocity (in both magnitude and sense, i.e., positive and negative), is

$$f_{R,\text{min}} = 2(V_{T,\text{max,closing}} + V_{T,\text{max,opening}} + V_g)/\lambda \tag{4.1}$$

where V_g is the upper limit for ground moving target rejection. V refers to the speed, or the magnitude of the range rate.

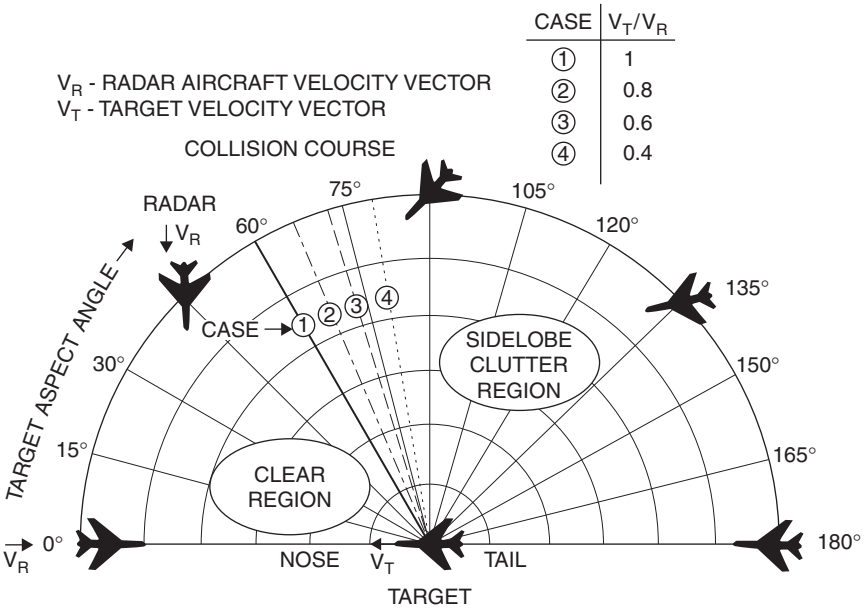


FIGURE 4.4 Sidelobe clutter-clear regions versus target aspect angle. Note the target is at the center of the plot with the radar platform on the circumference.

However, some pulse doppler radars employ a PRF that is unambiguous in velocity magnitude only, i.e., $f_{R,\min} = 2\,[\max(V_{T,\max,\text{closing}}, V_{T,\max,\text{opening}}) + V_g]/\lambda$, and rely on detections in multiple PRFs during the time on target to resolve the sign ambiguity in doppler. These radars can be described as *high-medium*-PRF and can be considered to be in the high-PRF category if the older definition of high PRF (no velocity ambiguity) is extended to allow one velocity ambiguity, that of doppler sense. The lower PRF eases the measurement of true range while retaining the high-PRF advantage of a single blind-speed region near zero doppler. High-medium PRF is becoming more prevalent in modern airborne radars for air-to-air search.

The choice between high and medium PRF involves a number of considerations, such as transmitter duty cycle limit, pulse compression availability, signal-processing capability, measurement accuracy requirements, etc., but often depends on the need for all-aspect target detectability. All-aspect coverage requires good performance in tail chase, where the target doppler is in the sidelobe clutter region near the altitude-line. In a high-PRF radar, the range foldover may leave little clear region in the range dimension, thus degrading target detectability. By using a lower or medium PRF, the clear region in range is increased at the expense of velocity foldover for high-doppler targets that are in the clutter-free region in high PRF. As an example, Figure 4.5 shows the clutter-plus-noise-to-noise ratio in range-doppler coordinates for two different X-band waveforms at similar altitudes and aircraft velocities. The range dimension represents the unambiguous range interval R_u, and the frequency dimension represents the PRF interval, with the main-beam clutter, altitude-line, and sidelobe clutter regions clearly discernible. In both waveforms, the main-beam clutter return is positioned to DC through clutter positioning via an

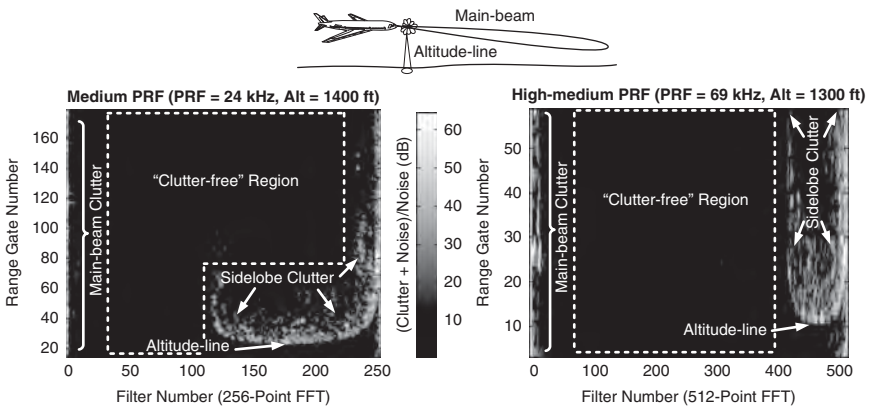


FIGURE 4.5 Clutter-plus-noise-to-noise ratio in range-doppler space

offset applied to the transmit frequency. The medium-PRF spectrum (PRF = 24 kHz) contains a range-doppler region in which the sidelobe clutter is below thermal noise and in which good tail-aspect target detectability can be achieved. The 69 kHz high-medium PRF waveform has a much more severe clutter folding, and tail aspect targets would compete with sidelobe clutter at nearly all ranges, but the clutter-free region is much larger.

Because the clutter is folded in both range and doppler with medium-PRF, a number of PRFs may be required to obtain a satisfactory probability of sufficient detections to resolve the range and doppler ambiguities. The multiple PRFs move the relative location of the clear regions so that all-aspect target coverage is achieved. Since the sidelobe clutter generally covers the doppler region of interest, the ratio of the region with sidelobe clutter below noise relative to the total range-doppler space is a function of the radar altitude, speed, and antenna sidelobe level.

If a high-PRF waveform is used, the clear-range region disappears because the sidelobe clutter folds in range into the unambiguous range interval (assuming the target doppler is such that it still competes with the sidelobe clutter). However, in those doppler regions free of sidelobe clutter, as shown in Figure 4.3 and Figure 4.4, target detectability is limited only by thermal noise, independent of radar altitude, speed, and sidelobe level. This requires system stability sidebands to be well below noise for the worst-case main-beam clutter. Thus, although medium PRF provides all-aspect target coverage, the target is potentially competing with sidelobe clutter at all aspects, whereas with high PRF, a target can become clear of sidelobe clutter at aspect angles forward of the beam aspect.

For targets with sufficient radial velocity, high PRF is typically more efficient than medium PRF. The transmit pulse width is usually limited by the transmitter's ability to preserve the pulse amplitude and phase characteristics over the duration of the transmit pulse. For a fixed transmit pulse width and peak power, a waveform with a higher PRF will have a higher transmit duty cycle resulting in a higher average transmit power. For a given coherent processing time, more energy is placed on the target, which improves detectability. For this reason, high PRF is used for long-range search of high-speed closing targets.

Range Gating. Range gating divides the time between transmit pulses into multiple cells or range gates. Range gating eliminates excess receiver noise and clutter from competing with the signal and permits target tracking and range measurement. The range gate is typically matched to the bandwidth of the transmit pulse. In a surveillance radar, a number of receiver gates are used to detect targets that may appear at any range within the interpulse period. Figure 4.6 illustrates the general case where the gate spacing τ_s, the gate width τ_g, and the transmitted pulse τ_t are all unequal. Selecting $\tau_t = \tau_g$ maximizes target return signal-to-noise ratio and, as a result, range performance. Selecting $\tau_g > \tau_s$ creates overlapped range gates and reduces the range gate straddle loss (Section 4.6) but can increase the possibility of range ghosts unless contiguous detections from straddled target returns are "clumped" prior to the ambiguity resolution (Section 4.4). With range gating, the range measurement accuracy is on the order of the range gate size (150 m/μs), but this can be improved to a fraction of the gate width by amplitude centroiding.

Timeline Definitions. Pulse doppler radar works on several different time scales. Various organizations have their own nomenclature for time-based parameters. Therefore, the timeline definitions used throughout this chapter are defined here.

Figure 4.7 illustrates the different time scales. Starting at the lowest level, a series of coherent pulses are transmitted at a pulse repetition frequency (PRF). The time between the pulses is the *interpulse period* (IPP), which is simply the inverse of the PRF. The receive portion of the IPP is broken up into range gates. The *transmit duty cycle* is the transmit pulse width divided by the IPP. The train of pulses is called the *coherent processing interval* (CPI). The coherent processing forms a bank of doppler

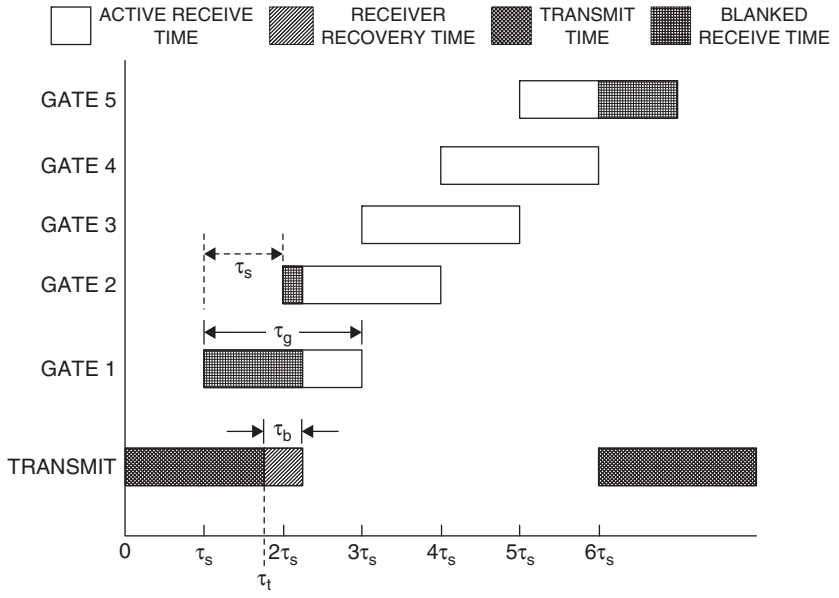


FIGURE 4.6 Example of range gates with 50% overlap equally spaced in the interpulse period. τ_b represents the extra blanking time after the transmit pulse to allow for receiver/protector recovery.

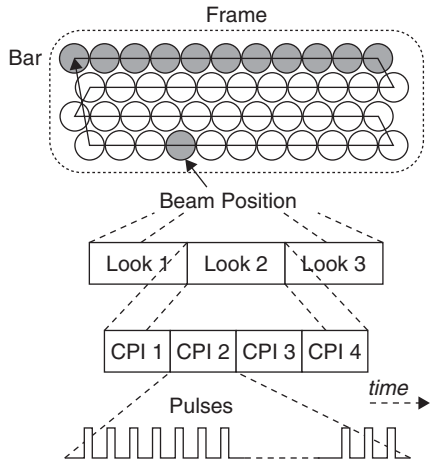


FIGURE 4.7 Pulse doppler dwell timeline

filters for each range gate resulting in a range-doppler map for a CPI, similar to that shown in Figure 4.5.

Several CPIs with the same PRF, but possibly different transmit carrier frequencies, can be noncoherently combined via *postdetection integration* (PDI). If frequency modulation (FM) ranging is used, all the CPIs that are noncoherently integrated must have the same FM slope. The grouping of CPIs is a *look*. Detections are determined for the range-doppler cells in a look.

Multiple looks with different PRFs or frequency modulations are used to resolve range and/or doppler ambiguities. This group of looks is a dwell. A dwell is associated with a particular antenna line-of-sight or *beam position*. Target reports are generated for each dwell.

A *bar* refers to a line of beam positions at a constant elevation. In search, a multibar *raster* scans the beam over an assigned area or volume to create a *frame*. A frame may have multiple bars. Typically, the antenna will visit every beam position once during a search frame.

Basic Configuration. Figure 4.8 shows a representative configuration of a pulse doppler radar utilizing digital signal processing under the control of a mission processor. Included are the antenna, receiver/exciter, signal processor, and data processor. The radar's control processor receives inputs from the on-board systems, such as the inertial navigation system (INS), and operator controls via the mission processor, and performs as a master controller for the radar hardware.

Coherent processing requires that all frequency down-conversions, including the final conversion to baseband, retain the coherent phase relationship between transmitted and received pulses. All the local oscillators are phase referenced to the same *master oscillator*, which is also used to produce the transmitted waveform. The *in-phase* (I) and *quadrature* (Q) components at baseband represent the real and imaginary parts, respectively, of a complex number whose complex argument in phasor notation is the phase difference between the transmitted and received pulses. The complex modulus, or magnitude, is proportional to the received echo strength.

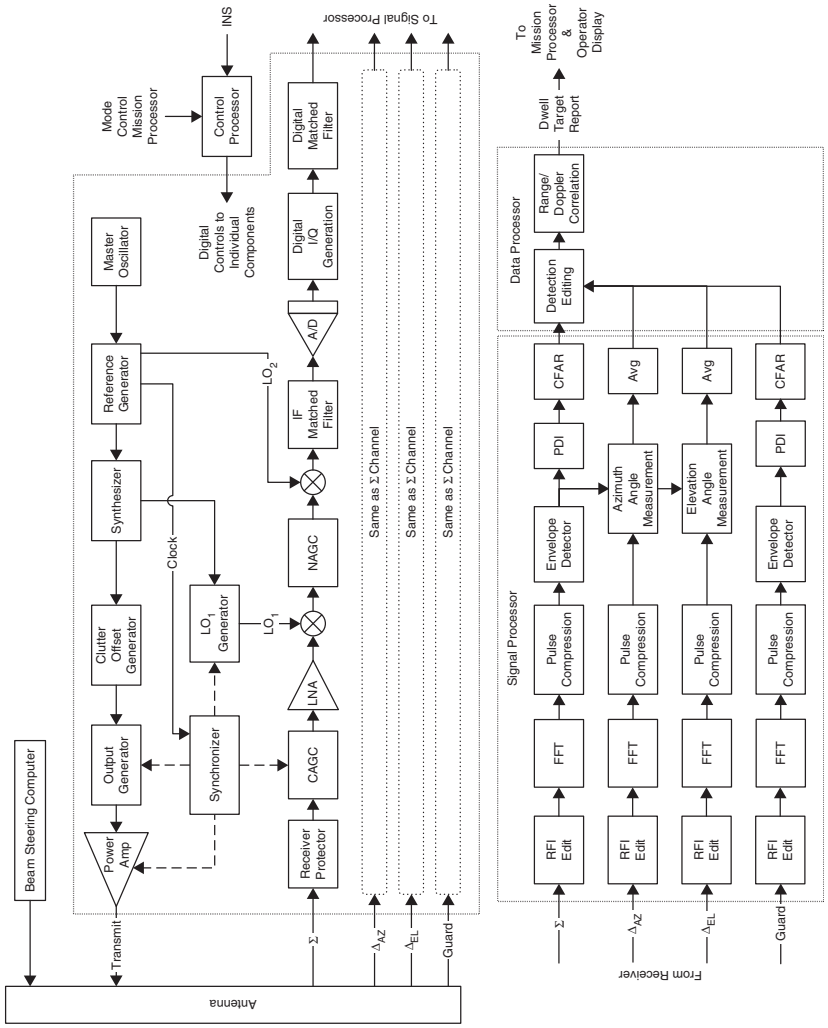


FIGURE 4.8 Typical pulse doppler radar configuration

Master Oscillator. The master oscillator provides a free-running, stable reference sinusoid on which the system synchronization is based.

Synchronizer. The synchronizer distributes precisely timed strobes and clocks for the various components of the radar system to ensure the time alignment of transmit waveforms and the reception of their corresponding returns. These low-jitter timing signals are used to enable and disable the transmit power amplifier to create the transmit pulse train, blank the receiver during transmission, and form the range gates.

Reference Generator. The reference generator outputs fixed frequency clocks and local oscillators (LOs).

Synthesizer. The synthesizer generates the transmit carrier frequency and the first local oscillator (LO_1) frequency. Frequency agility is provided to the transmit and LO_1 signals.

Clutter Offset Generator. The clutter offset generator shifts the transmit carrier slightly, so that on receive the main-beam clutter is positioned at zero doppler frequency, or DC (direct current) after basebanding. The same effect could be obtained by shifting the receiver LO_1 frequency. With the clutter at DC, the spurious signals caused by certain receiver nonlinearities, such as mixer intermodulation products and video harmonics, also fall near DC and can be filtered out along with the main-beam clutter.[19] The frequency shift applied is a function of the antenna main-beam line-of-sight relative to the platform's velocity vector. This process is known as *clutter positioning*.

Output Generator. The output generates the pulsed radio frequency (RF) transmit signal, which is the transmit drive signal that is amplified by the power amplifier prior to being fed to the transmit antenna.

Antenna. The antenna can be mechanically or electronically scanned. Modern pulse doppler radars have migrated to the use of active electronically scanned arrays (AESAs).[20] AESAs contain transmit/receive (T/R) modules, each comprising a transmit power amplifier and a receive low-noise amplifier (LNA) along with an attenuator and phase shifter, at each antenna element.

If the same antenna is used for transmit and receive, a duplexer must be included. This duplexer is usually a passive device, such as a circulator, which effectively switches the antenna between the transmitter and receiver. Considerable power may be coupled to the receiver since typically no more than 20 to 25 dB of isolation may be expected from ferrite circulators.

Antennas may form various beams. The transmit beam can be formed with uniform aperture illumination to maximize the amount of energy on target, whereas the receive sum (Σ) beam is typically formed with a low-sidelobe taper to minimize the returns from ground clutter. The Σ beam is used for target detection and, acting as a spatial filter, is the first line of defense against clutter and interference in the sidelobe region. To facilitate target tracking, angle measurements with accuracies finer than the antenna beamwidth are usually required. A technique to obtain such angle measurements of a target on a single pulse is called *monopulse*. Monopulse can be characterized as amplitude or phase, with phase being preferable due to its advantage in angle accuracy for a given signal-to-noise ratio. Phase monopulse uses a *delta* or difference beam,

which is essentially formed by dividing the aperture into two halves and subtracting the corresponding phase centers. Monopulse beams, delta-azimuth (Δ_{AZ}) and delta-elevation (Δ_{EL}), are formed to provide phase monopulse azimuth and elevation angle measurements.[21] Self-calibration routines controlled by the control processor ensure that the phase and amplitude match of the receiver channels enables accurate monopulse measurements. A guard beam with a near-omnidirectional pattern is formed for sidelobe detection blanking as discussed in Section 4.2.

Receiver/Protector (R/P). The receiver/protector is a low-loss, fast-response, high-power switch that prevents the transmitter output from the antenna's duplexer from damaging the sensitive receiver front end. Fast recovery is required to minimize desensitization in the range gates following the transmitted pulse. R/Ps can be implemented with a gas discharge tube, in which a gas is ionized by high-power RF. A diode limiter can be used instead of or in conjunction with the gas discharge tube. The R/P can be reflective or absorptive, but must have low insertion loss to minimize impact on receive chain noise figure.

Clutter Automatic Gain Control (CAGC). The CAGC attenuator is used both for suppressing transmitter leakage from the R/P into the receiver (so the receiver is not driven into saturation, which could lengthen recovery time after the transmitter is turned off) and for controlling the input signal levels into the receiver. The received levels are kept below saturation levels, typically with a clutter AGC in search and a target AGC in single-target track, to prevent spurious signals, which degrade performance, from being generated.

Noise Automatic Gain Control (NAGC). The NAGC attenuator is used to set the thermal noise level in the receiver to support the required dynamic range, as discussed in Section 4.3. The attenuation is commanded based on measurements of the noise during periodic calibration.

Digital Preprocessing. The advent of high-speed, high-dynamic range analog-to-digital converters (A/Ds) allows IF-sampling and digital basebanding. The digital IF-sampled output of the receiver is downconverted to baseband (DC) via a digital product detector (DPD).[22] Superior I/Q image rejection is an advantage of a DPD.

The I and Q signals are passed through the digital portion of the pulse matched filter. The combination of the IF matched filter and the digital matched filter form the receiver's single-pulse matched filter.

Digital Signal Processing. Following digital preprocessing is a doppler filter bank for main-beam clutter rejection and coherent integration. RF interference (RFI) that is pulsed and asynchronous to the radar timing can often be detected prior to the coherent integration. Range-IPP cells where RFI is detected are then "repaired" to prevent corruption of the output spectrum. The filter bank is usually realized by using the fast Fourier transform (FFT); however, the discrete Fourier transform (DFT) can be used when the number of filters is small. Appropriate weighting is employed to reduce the filter sidelobes. The amount of weighting can be chosen adaptively by sensing the peak signal levels (usually main-beam clutter) and selecting the doppler weighting dynamically.

If pulse compression modulation is used on the transmit pulse to increase energy on target, pulse compression can be performed digitally either before or after the doppler

filter bank. The advantage of pulse compression after the filter bank is that the effects of doppler on pulse compression can be largely removed by tailoring the pulse compression to the doppler offset of each doppler filter. However, this increases the total amount of signal processing required.

The envelope at the output of the FFT is formed with a linear ($\sqrt{I^2 + Q^2}$) or square-law ($I^2 + Q^2$) detector. Historically, linear detectors were used to manage dynamic range in fixed-point processors. Square-law detectors are preferred for some modern floating-point processors. Postdetection integration (PDI) may be used where each range-gate–doppler-filter output is linearly summed over several CPIs. For each range-doppler cell in the Σ channel, the PDI output is compared with a detection threshold determined by a constant-false-alarm-rate (CFAR) process.[23–26] Cells with amplitudes greater than the CFAR threshold are labeled as detections.

Similar processing is done in the Δ_{AZ} and Δ_{EL} channels with exceptions, as shown in Figure 4.8. For those range-doppler cells with declared detections, the imaginary part of the Δ_{AZ}/Σ and Δ_{EL}/Σ ratios are used for phase comparison monopulse to estimate the azimuth and elevation angles, respectively, relative to the center of the Σ main beam. The angle estimates are computed for each coherent look and then averaged over the number of CPIs noncoherently integrated via PDI.

The guard channel is processed similar to the Σ channel. The guard channel's purpose is to blank sidelobe detections, as described in Section 4.2.

Postprocessing. Following the CFAR is detection editing, which contains the sidelobe discrete rejection logic. Following detection editing, range and velocity ambiguity resolvers work over several looks within a dwell. The final detection outputs, along with their corresponding unambiguous range, velocity, and angle measurements, and their estimated accuracies, are passed to the mission processor for tracking and operator display.

4.2 PULSE DOPPLER CLUTTER

General. Clutter returns from various scatterers have a strong influence on the design of a pulse doppler radar as well as an effect on the probability of detection of point targets. Clutter scatterers include terrain (both land and sea), weather (rain, snow, etc.), and chaff. Since the antennas generally used in pulse doppler radars have a single, relatively high-gain main beam, main-beam clutter may be the largest signal handled by the radar when in a down-look condition. The narrow beam limits the frequency extent of this clutter to a relatively small portion of the doppler spectrum. The remainder of the antenna pattern consists of sidelobes, which result in sidelobe clutter. This clutter is generally much smaller than the main-beam clutter but covers much more of the frequency domain. The sidelobe clutter from the ground directly below the radar, the altitude-line, is frequently large owing to a high reflection coefficient at steep grazing angles, the large geometric area, and the short range. Range performance is degraded for targets in the sidelobe clutter region wherever the clutter is near or above the receiver noise level. Multiple PRFs may be used to move the target with respect to the sidelobe clutter in the range-doppler map, thus avoiding completely blind ranges or blind frequencies due to high clutter levels. This relative motion occurs due to the range and doppler foldover from range and/or doppler ambiguities. If one PRF folds sidelobe clutter and a target to the same apparent range and doppler, a sufficient change of PRF will separate them.

Ground Clutter in a Stationary Radar. When the radar is fixed with respect to the ground, both stationary main-beam and sidelobe clutter returns occur at zero-doppler offset from the transmit carrier frequency. The sidelobe clutter is usually small compared with main-beam clutter as long as some part of the main beam strikes the ground. The clutter can be calculated as in a pulsed radar, then folded in range as a function of the PRF.

Ground Clutter in a Moving Radar. When the radar is moving with a velocity V_R, the clutter is spread over the frequency domain as illustrated in Figure 4.2 for the special case of horizontal motion. The foldover in range and doppler is illustrated in Figure 4.9 for a medium-PRF radar where the clutter is ambiguous in both range and doppler. The radar platform is moving to the right at 1000 kt with a dive angle of 10°. The narrow annuli (iso-range contours) define the ground area that contributes to clutter in the selected range gate. The five narrow hyperbolic bands (iso-doppler contours) define the area that contributes to clutter in the selected doppler filter. The shaded intersections represent the area, or clutter patches, that contributes to the range-gate–doppler-filter cell. Each clutter patch contributes clutter power as a function of the antenna gain in the direction of the clutter patch and the reflectivity of the clutter patch.

The main beam illuminates the elliptical area to the left of the ground track. Since this area lies entirely within the filter area, the main-beam clutter falls within this filter, and all other filters receive sidelobe clutter. Four range annuli are intersected by the main-beam ellipse, so the main-beam clutter in this range gate is the vector sum of the signals received from all four clutter patches. Owing to this high degree of range foldover, all range gates will have approximately equal clutter.

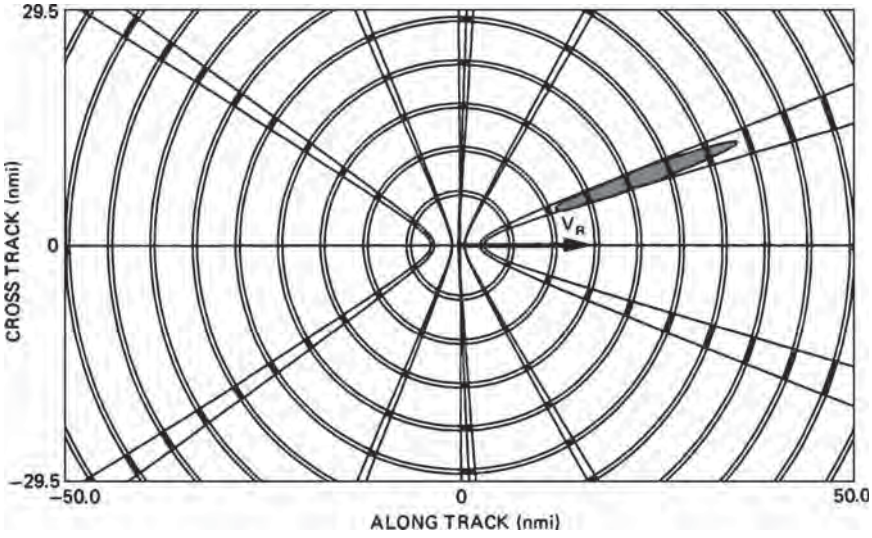


FIGURE 4.9 Plan view of range-gate and doppler-filter areas. Radar altitude = 10,000 ft; velocity = 1000 kt to right; dive angle = 10°; radar wavelength = 3 cm; PRF = 15 kHz; range gate width = 6.67 μs; range gate = 4; doppler filter at 2 kHz; bandwidth = 1 kHz; beamwidth = 5° (circular); main-beam azimuth = 20°; depression angle = 5°.

If the main beam were scanned 360° in azimuth with the same radar platform kinematics, the main-beam clutter would scan in doppler frequency so that it would appear in the selected filter ten times (twice for each hyperbolic band). In between, the filter would receive sidelobe clutter from all darkened intersections. With the use of the proper clutter offset (which would vary as a function of main-beam azimuth) on the transmit frequency, as described in Section 4.1, the doppler of the main-beam clutter return will be zero or DC.

Clutter Return: General Equations. The clutter-to-noise ratio from a single clutter patch with incremental area dA at a range R is

$$C/N = \frac{P_{\text{av}} G_T G_R \lambda^2 \sigma^0 dA}{(4\pi)^3 R^4 L_C k T_s B_n} \tag{4.2}$$

where P_{av} = average transmit power
G_T = transmit gain in patch direction
G_R = receive gain in patch direction
λ = operating wavelength
σ^0 = clutter backscatter coefficient
L_C = losses applicable to clutter
k = Boltzmann's constant = 1.38054×10^{-23} W/(Hz/K)
T_s = system noise temperature, K
B_n = doppler filter bandwidth

L_C refers to losses that apply to distributed surface clutter, as opposed to discrete, resolvable targets. These losses will be discussed in Section 4.6.

The clutter-to-noise ratio from each radar resolution cell is the integral of Eq. 4.2 over the doppler and range extent of each of the ambiguous cell positions on the ground.[27–31] Under certain simplified conditions, the integration can be closed-form,[32] but in general, numeric integration is required.

Main-beam Clutter. The net main-beam clutter-to-noise power in a single range gate in the receiver can be approximated from Eq. 4.2 by substituting the range gate's intersected area ($\frac{c\tau}{2\cos(\alpha)} R\theta_{\text{az}}$) within the main beam on the ground for dA and summing over all ambiguities of that range gate that are within the main beam.[33]

$$\frac{C}{N} = \frac{P_{\text{av}} \lambda^2 \theta_{\text{az}} (c\tau/2)}{(4\pi)^3 L_C k T_s B_n} \sum \frac{G_T G_R \sigma^0}{R^3 \cos(\alpha)} \tag{4.3}$$

The summation limits are the lower and upper edges in the elevation dimension of the smaller of the transmit and receive beams

where θ_{az} = azimuth half-power beamwidth, radians
τ = compressed pulse width
α = grazing angle at clutter patch

The remaining terms are as defined following Eq. 4.2.

If the main beam is pointed below the horizon, the main-beam clutter spectral width Δf due to platform motion measured 6 dB down from the peak is approximately[34]

$$\Delta f = \frac{2V_R}{\lambda} \left\{ \theta_B \cos(\phi_0) \sin(\theta_0) + \frac{\theta_B^2 \cos(\phi_0) \cos(\theta_0)}{8} + \frac{c\tau \sin^3(\phi_0) \cos(\theta_0)}{2h \cos(\phi_0)} \right\} \tag{4.4}$$

where V_R = radar ground speed
λ = RF wavelength
θ_B = 3 dB one-way antenna azimuth beamwidth, radians
ϕ_0 = main-beam depression angle relative to local horizontal, radians
θ_0 = main-beam azimuth angle relative to the horizontal velocity, radians
τ = compressed pulse width
h = radar altitude

When the magnitude of the main-beam azimuth angle is greater than half of the azimuth beamwidth ($|\theta_0| \geq \theta_B/2$), the main-beam clutter power spectral density can be modeled with a gaussian shape with a standard deviation $\sigma_c = 0.3\Delta f$.

Main-beam Clutter Filtering. In a pulse doppler radar utilizing digital signal processing, main-beam clutter is rejected by either a combination of a delay-line clutter canceler (MTI filter) followed by a doppler filter bank or by a filter bank with low filter sidelobes, which are achieved via weighting.[35] In either case, the filters around the main-beam clutter are blanked to minimize false alarms on main-beam clutter. This blanked region in doppler is known as the main-beam clutter *notch*.

The choice between these options is a trade-off of quantization noise and complexity versus the filter-weighting loss. If a canceler is used, filter weighting can be relaxed over that with a filter bank alone, since the canceler reduces the dynamic-range requirements into the doppler filter bank (if the main-beam clutter is the largest signal). Without a canceler, heavier weighting is needed to reduce sidelobes to a level so that the filter response to main-beam clutter is below the thermal-noise level. This weighting increases the filter noise bandwidth and hence increases the loss in signal-to-noise ratio.

Choosing the proper weighting is a compromise between rejecting main-beam clutter and maximizing target signal-to-noise ratio. To dynamically make this compromise, the filter weighting can be adaptive to the main-beam clutter level by measuring the peak return level (usually main-beam clutter) over the IPPs, and selecting or computing the best weighting to apply across the CPI. Another technique that is applicable to high-medium and high PRF is to generate a hybrid filter weighting by convolving two weighting functions. The result is a filter with significantly less weighting loss and low far-out sidelobes, but at a cost of relatively high near-in sidelobes.

To evaluate the effect of main-beam clutter on target detection performance, the clutter-to-noise ratio must be known for each filter where targets are to be detected. A general measure that can be easily applied to specific clutter levels is the improvement factor I. When using a doppler filter bank, as opposed to an MTI filter, the improvement factor is defined for each doppler filter as the ratio of the signal-to-clutter power at the output of the doppler filter to the signal-to-clutter power at the input.[36] The signal is assumed to be at the center of the doppler filter. Incorporating the effect of filter weighting, the improvement factor for a doppler filter is given by[37]

$$I(K) = \frac{\left[\sum_{n=0}^{N-1} A_n\right]^2}{\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} A_n A_m \exp\left\{-2\left[\pi(n-m)\sigma_c T\right]^2\right\}\cos\left[2\pi K(n-m)/N\right]} \tag{4.5}$$

where A_i = IPP weight, $0 \le i \le N-1$
N = number of IPPs in CPI
σ_c = standard deviation of clutter spectrum
K = filter number ($K = 0$ is the DC filter)
T = interpulse period

Clutter-transient Suppression. When (1) the PRF is changed for multiple-PRF ranging, (2) the slope is changed in linear FM ranging, or (3) the RF carrier is changed, the transient change in the clutter return may cause degradation unless it is properly handled.[38] Since the clutter is usually ambiguous in range in a pulse doppler radar, the clutter power increases at each interpulse period (IPP) as clutter return is received from the farther ambiguities, until the horizon is reached. This phenomenon is called *space charging*. Note that although an increasing number of clutter returns are received during the charging period, the vector sum may actually decrease owing to the random phase relations of the returns from different patches.

If a clutter canceler (MTI filter) is used, the output cannot begin to settle to its steady-state value until space charging is complete. Some settling time must be allowed before signals are passed to the filter bank. Therefore, the coherent integration time available during each CPI is reduced from the total CPI time by the sum of the space charge time and the transient settling time. The canceler settling time can be eliminated by *precharging* the canceler with the steady-state input value.[39] This is done by changing the canceler gains so that all delay lines achieve their steady-state values on the first IPP of data.

If no canceler is used, signals can be passed to the filter bank after the space charge is complete, so that the coherent integration time is the total CPI time minus the space charge time.

Altitude-line Clutter Blanking. The reflection from the earth directly beneath an airborne pulse radar is called altitude-line clutter. Because of specular reflection over smooth terrain, the large geometric area, and relatively short range, this signal can be large. It lies within the sidelobe clutter region of the pulse doppler spectrum.

Because it can be much larger than diffuse sidelobe clutter and usually has a relatively narrow spectral width, altitude-line clutter is often removed either by a special CFAR that prevents detection of the altitude-line, or by a tracker-blanker that removes these reports from the final output. In the case of the tracker-blanker, a closed-loop tracker is used to position range and velocity gates around the altitude return and blank the affected range-doppler region. Note that at very low altitudes, the angles that subtend the first range gate on the ground can be quite big, and the spectral width widens.

Sidelobe Clutter. The entire clutter spectrum can be calculated for each range gate by Eq. 4.2 if the antenna pattern is known in the lower hemisphere. In preliminary system design, the exact gain function may not be known, so one useful approximation is that the sidelobe radiation is isotropic with a constant gain of G_{SL}.

Sidelobe Discretes. An inherent characteristic of airborne pulse doppler radars is that echoes from large, resolvable objects on the ground (discretes), such as buildings, may be received through the antenna sidelobes and appear as though they were

smaller moving targets in the main beam. This is a particularly severe problem in a medium-PRF radar, where all-aspect target performance is usually desired, since these returns compete with targets of interest. In a high-PRF radar, there is little if any range region clear of sidelobe clutter, such that the sidelobe clutter portion of the doppler spectrum is often not processed (since target detectability is severely degraded in this region). Further, in a high-PRF radar, especially at higher altitudes, the relative amplitudes of the distributed sidelobe clutter and the discrete returns are such that the discretes are not visible in the sidelobe clutter.

The apparent radar cross section (RCS), σ_{app}, of a sidelobe discrete with an RCS of σ is $\sigma_{app} = \sigma G_{SL}^2$, where G_{SL} is the sidelobe gain relative to the main beam. The larger-size discretes appear with a lower density than the smaller ones, and a model commonly assumed at the higher radar frequencies is shown in Table 4.4. Thus, as a practical matter, 10^6 m^2 discretes are rarely present, 10^5 m^2 are sometimes present, and 10^4 m^2 are often present.

Two mechanizations for detecting and eliminating false reports from sidelobe discretes are the guard channel and postdetection sensitivity time control (STC). These are discussed in the paragraphs that follow.

Guard Channel. The guard channel mechanization compares the outputs of two parallel receiving channels, one connected to the main antenna and the second to a guard antenna (the Σ and Guard channel in Figure 4.8, respectively), to determine whether a received signal is in the main beam or the sidelobes.[40–44] The guard channel uses a broad-beam antenna that (ideally) has a pattern above the main-antenna sidelobes. The returns from both channels are compared for each range-doppler cell that had a detection in the main channel. For these range-doppler cells, when the guard channel return is greater than that of the main channel, the detection is rejected (blanked). If the main channel return is higher, the detection is passed on.

A block diagram of a guard channel mechanization is shown in Figure 4.10. After the CFAR (which ideally would be identical in both channels), there are three thresholds: the main channel, guard channel, and main-to-guard-ratio thresholds. The detection logic of these thresholds is also shown in Figure 4.10.

The blanking that occurs because of the main/guard comparison affects the detectability in the main channel, the extent of which is a function of the threshold settings. The threshold settings are a tradeoff between false alarms due to sidelobe returns and detectability loss in the main channel. An example is shown in Figure 4.11 for a nonfluctuating target, where the ordinate is the probability of detection in the final output of the sidelobe blanker and the abscissa is the signal-to-noise ratio (SNR) in the main channel. The quantity B^2 is the ratio of the guard channel SNR to the main channel SNR and is illustrated in Figure 4.12.

TABLE 4.4 Discrete Clutter Model

Radar Cross Section (m^2)	Density (per square mile)
10^6	0.01
10^5	0.1
10^4	1

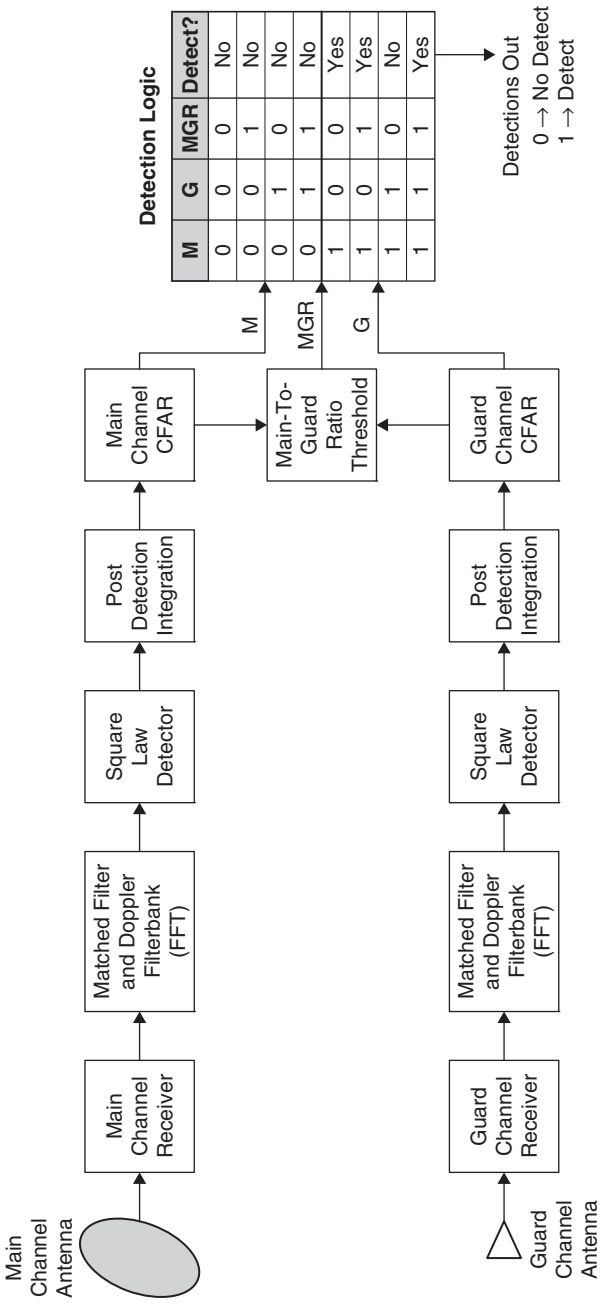


FIGURE 4.10 Two-channel sidelobe blanker

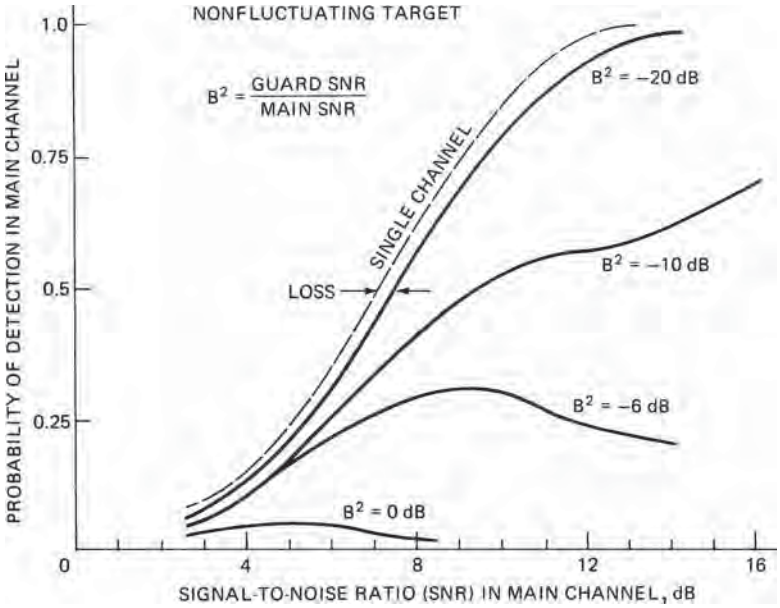


FIGURE 4.11 Probability of detection versus signal-to-noise ratio with a guard channel

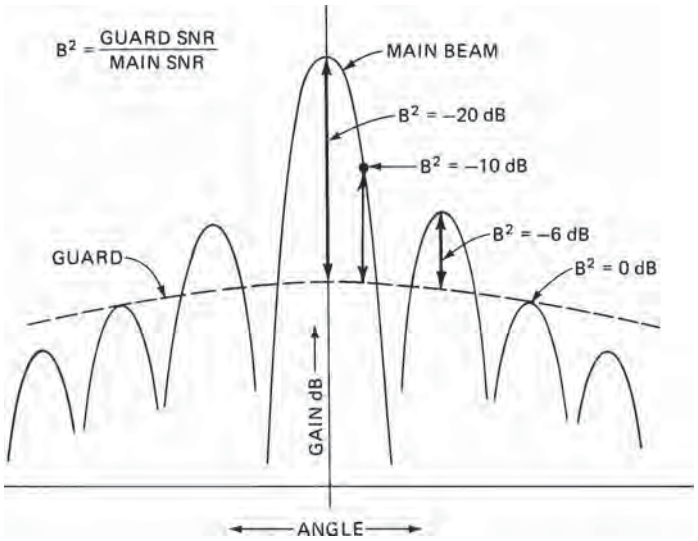


FIGURE 4.12 Main and guard antenna patterns

B^2 is small for a target in the main beam and large, 0 dB or so, for a target at the sidelobe peaks. In the example shown, there is a 0.5 dB detectability loss due to the guard blanking for targets in the main beam.

Ideally, the guard antenna gain pattern exceeds that of the main antenna at all angles in space (except for the main beam) to minimize detections through the sidelobes. If not, however, as illustrated in Figure 4.11 and Figure 4.12, returns through the sidelobe peaks of the main pattern above the guard pattern have a significant probability of detection in the main channel and would represent false detections.

Postdetection STC. In the ambiguity resolution, as the output returns are range-correlated, they are subjected to postdetection STC, or RCS thresholding, applied inside the range correlation process. Target returns that range correlate inside the STC range, but fall below the STC threshold, are likely sidelobe discretes and are blanked or removed from the correlation process (and kept from ghosting with other targets).

The basic logic is shown in Figure 4.13.[45] Basically, the CFAR output data is correlated (resolved) in range three times. Each correlator calculates unambiguous range using *M* out of the *N* sets of detection data (e.g., three detections required out of eight PRFs). No doppler correlation is used since the doppler is ambiguous. The results of the first two correlations are used to blank all outputs that are likely to be sidelobe discretes from the final range correlator. Here, three range correlators are used in which the first, the *A* correlator, resolves the range ambiguities within some nominal range, say, 10 nm, beyond which sidelobe discretes are not likely to be detected. A second correlator, the *B* correlator, resolves the range ambiguities out to the same range, but before a target can enter the *B* correlator, its amplitude is thresholded by a range-varying threshold (the STC threshold). A range gate by range gate comparison is made of the correlations in the *A* and *B* correlators, and if a range gate correlates in A and not in B, that gate is blanked out of the third correlator, the *C* correlator. The *C* correlator resolves the range ambiguities within the maximum range of interest. An alternative mechanization is to replace the range-varying STC with an equivalent RCS threshold inside the range correlation process. The RCS is computed for each possible unfolded range (starting from the shortest range) and compared to the RCS threshold. Detections that range correlate, but are below the RCS threshold, are prevented from correlating with other detects (and all of their unfolded ranges are also prevented from correlating).

The principle behind the postdetection STC approach is illustrated in Figure 4.14, where the return of a target in the main beam and a large discrete target in the sidelobes is plotted versus unambiguous range (that is, after the range ambiguities have been resolved). Also shown are the normal CFAR threshold and the STC threshold versus range. A discrete return in the sidelobes is below the STC threshold, and a return in the main beam is above the threshold, such that the sidelobe discrete can be recognized and blanked without blanking the target in the main beam. The STC onset range represents the range at which a large discrete target in the sidelobes exceeds the CFAR threshold.

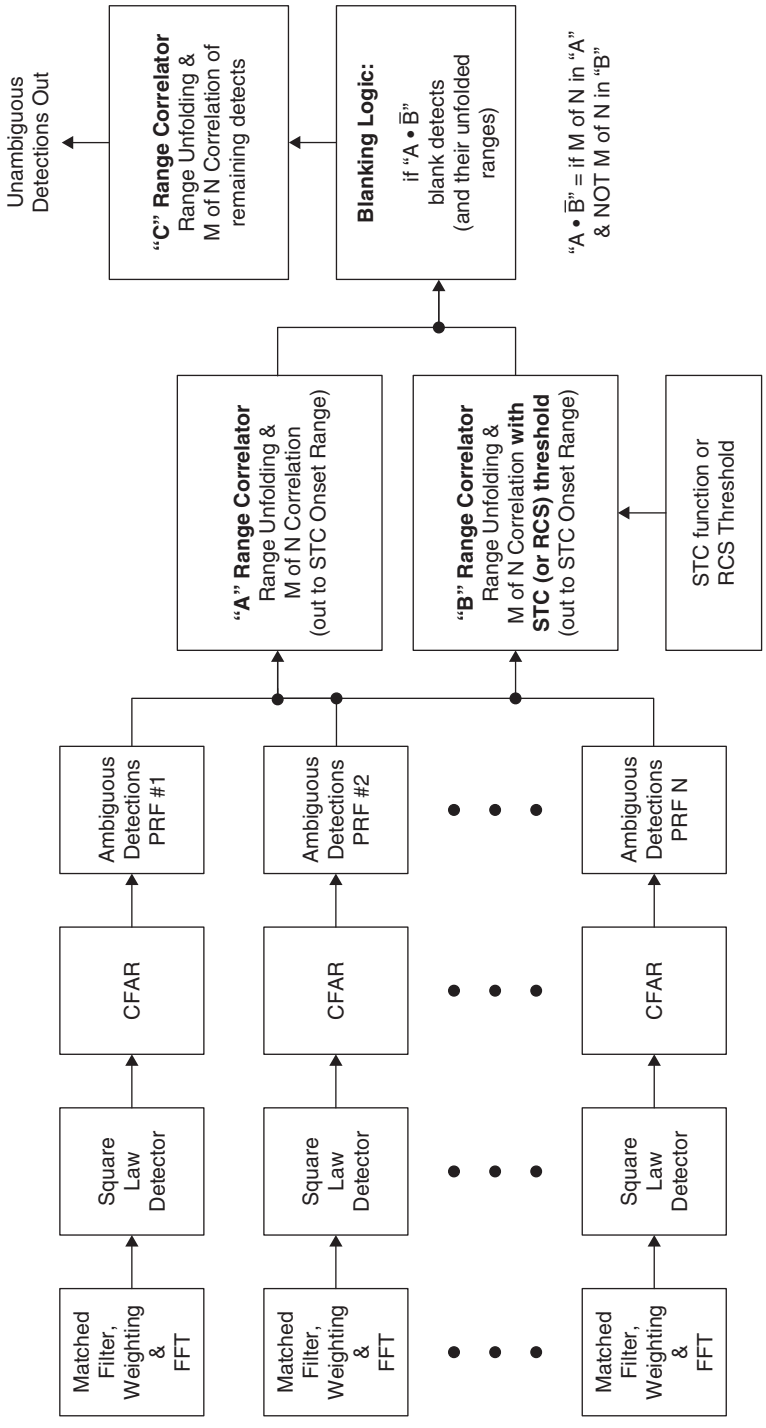

Unambiguous Detections Out
"C" Range Correlator
Range Unfolding & M of N Correlation of remaining detects
Blanking Logic:
if "A • B̄" blank detects (and their unfolded ranges)
"A • B̄" = if M of N in "A" & NOT M of N in "B"
"A" Range Correlator
Range Unfolding & M of N Correlation (out to STC Onset Range)
"B" Range Correlator
Range Unfolding & M of N Correlation with STC (or RCS) threshold (out to STC Onset Range)
STC function or RCS Threshold
Ambiguous Detections PRF #1
Ambiguous Detections PRF #2
Ambiguous Detections PRF N
CFAR
CFAR
CFAR
Square Law Detector
Square Law Detector
Square Law Detector
Matched Filter, Weighting & FFT
Matched Filter, Weighting & FFT
Matched Filter, Weighting & FFT

FIGURE 4.13 Single-channel sidelobe blanker using postdetection STC or RCS thresholding to remove sidelobe discretes.

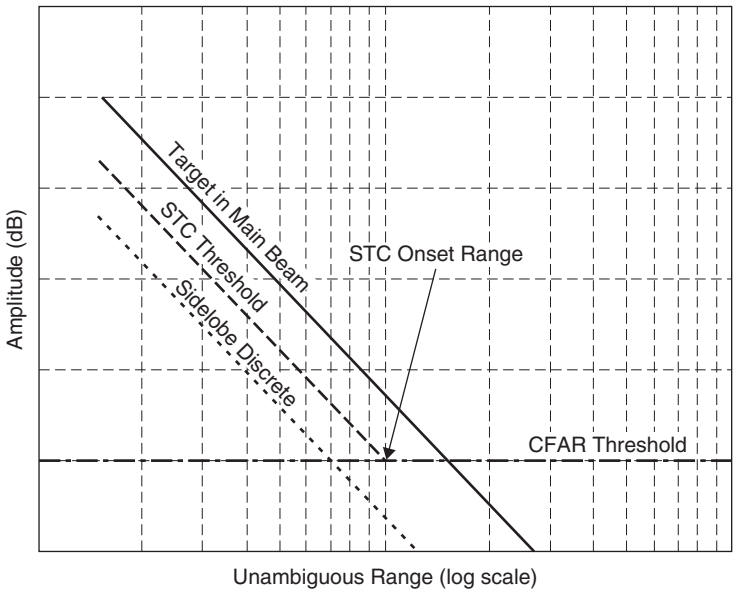


FIGURE 4.14 Postdetection STC levels

4.3 DYNAMIC-RANGE AND STABILITY REQUIREMENTS

Doppler processing separates moving targets from clutter and allows them to be detected while only competing against thermal noise, assuming that the targets have sufficient radial velocity ($> 2V_R/\lambda$) and the PRF is high enough for an unambiguous clutter spectrum. Coherence, the consistency of phase of a signal's carrier frequency from one pulse to the next, is crucial for doppler processing. Without careful system design, amplitude and phase instabilities during the coherent integration time will broaden the main-beam clutter spectrum and raise the noise floor that clutter-free targets must compete with for detection. Nonlinearities in the system can also cause discrete spurious spectral signals that can be mistaken as targets. The instantaneous dynamic range of the system governs the system linearity and hence sensitivity in a strong clutter environment. The driving factor upon stability requirements is when the main-beam clutter level is at the saturation point of the receiver.

Dynamic Range. Dynamic range, as discussed here, can be referred to as *instantaneous dynamic range* and is the linear region above thermal noise over which the receiver and signal processor operate before any saturation (clipping) or gain limiting occurs. If saturations occur, spurious signals that degrade performance may be generated. For example, if main-beam clutter saturates, spurious frequencies can appear in the doppler passband normally clear of main-beam clutter, and this may generate false-target reports. An automatic gain control (AGC) function is often employed to prevent saturations on either main-beam clutter in search or the target in Single-Target Track mode. However, the use of AGC degrades the system's sensitivity, so large

instantaneous dynamic range is preferable. If saturations occur in a range gate during an integration period, an option in a multiple-range gated system is simply to blank detection reports from that gate. When a MTI filter is not used, the doppler filter bank for each range gate can be examined to determine if there are any detections due to spurious signals from large clutter, with subsequent editing of these detections if the measured clutter-to-noise ratio exceeds the dynamic range. Similar logic can be applied to saturated range gates to determine if the largest signal in the filter bank is in the passband or represents saturated clutter returns. Saturated returns with the peak signal in the doppler passband can represent valid targets at short ranges and need not be subjected to the sidelobe blanking logic.

The most stressing dynamic-range requirement is due to main-beam clutter when searching for a small, low-flying targets. Here, full sensitivity must be maintained in the presence of the clutter to maximize the probability of detecting the target.

The dynamic-range requirement of a pulse doppler radar, as determined by main-beam clutter, is a function not only of the basic radar parameters such as power, antenna gain, etc., but of radar altitude above the terrain and the radar cross section (RCS) of low-flying targets. As an example, Figure 4.15 shows the maximum clutter-to-noise ratio (C/N_{max}) that appears in the ambiguous-range interval, i.e., after range folding, for a medium-PRF radar as a function of radar altitude and the range of the intersection of the peak of the main-beam with the ground. Note that the clutter-to-noise ratio is a rms power ratio measured at the A/D converter. A peak power ratio would be 3 dB higher.

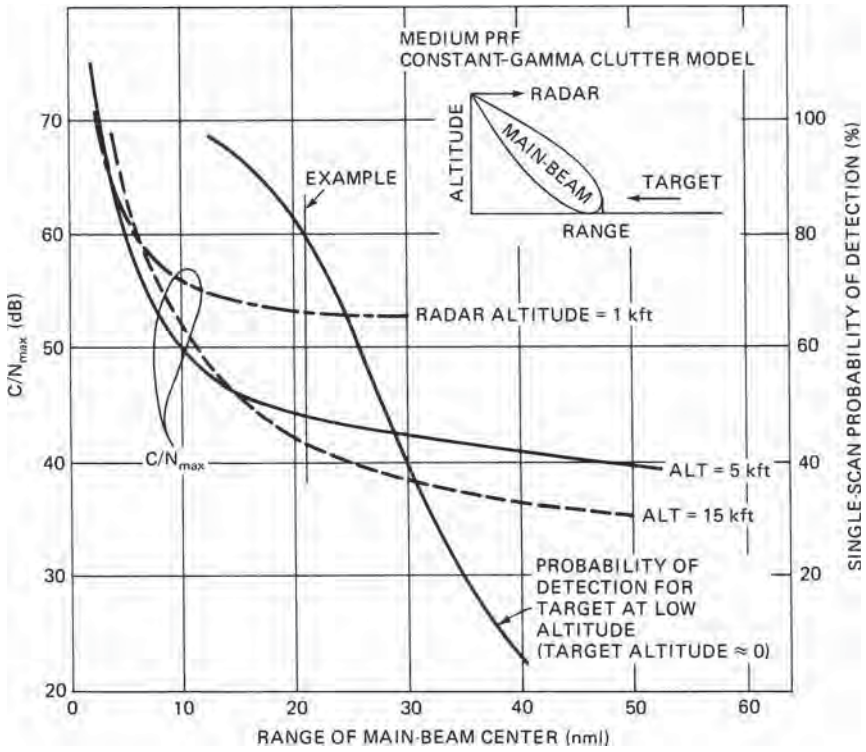


FIGURE 4.15 Dynamic-range example

The amplitude of clutter returns fluctuate over time and are modeled as a stochastic process. The clutter-to-noise ratio represents the mean value of this process over time. Figure 4.15 assumes a pencil-beam antenna pattern and a constant-gamma model for clutter reflectivity.[46] The antenna beam is pointed at the ground corresponding to the range of the target. At longer ranges (small look-down angles), clutter decreases with increasing radar altitude since range folding is less severe owing to less of the main beam intersecting the ground. At shorter ranges, clutter increases with radar altitude since the clutter patch size on the ground increases. While Figure 4.15 is for a medium-PRF radar, similar curves result for a high-PRF radar.

Also shown in Figure 4.15 is the single-scan probability of detection P_d versus range for a given RCS target in a receiver with unlimited dynamic range. If it is desired to have the low-flying target reach at least, say, an 80% P_d before any gain limiting (i.e., the use of AGC) occurs, the dynamic-range requirement is driven by the main-beam clutter levels $C/N_{\max}$ of 53 dB at 1000 ft, 44 dB at 5000 ft, and 41 dB at 15,000 ft for this example. The higher the desired probability of detection or the lower the radar altitude, the more dynamic range is required. Further, if the specified target RCS is reduced, the dynamic-range requirement for the same desired P_d increases as the P_d-versus-range curve in Figure 4.15 shifts to the left.

In a pulse doppler radar using digital signal processing, the A/D converters are usually selected to have a dynamic range that meets or exceeds the usable dynamic range set by the maximum clutter-to-noise ratio ($C/N_{\max}$) and the system stability. The peak dynamic range, defined as the maximum peak sinusoidal signal level relative to the rms thermal-noise level that can be processed linearly, is related to the number of amplitude bits in the A/D converter by

$$\left[\frac{S_{\max}}{N}\right]_{\text{dB}} = 20\log_{10}\left(\frac{2^{N_{\text{AD,amp}}}-1}{[\text{noise}]_{\text{quanta}}}\right) \tag{4.6}$$

where $[S_{\max}/N]_{\text{dB}}$ = maximum input peak sinusoidal level relative to rms noise, dB
$N_{\text{AD,amp}}$ = number of amplitude bits (not including sign bit) in the A/D converter
$[\text{noise}]_{\text{quanta}}$ = rms thermal-noise voltage level at the A/D converter, quanta

The rms thermal-noise voltage level at the A/D converter is given in terms of *quanta*. A single quanta refers to a unit quantization level of the A/D converter.

From the relationship described above and assuming the A/D converter limits the dynamic range, the A/D converter size can now be determined. Additional margin to allow for main-beam clutter fluctuations above the mean value also needs to be considered. Since main-beam clutter time fluctuation statistics are highly dependent on the type of clutter being observed, such as sea clutter or clutter from an urban area, and are generally unknown, a value of 10 to 12 dB above the rms value is often assumed for the maximum peak level (this also includes the 3 dB difference between the rms and peak values of a sinusoidal signal). Thus, the required number of amplitude bits in the A/D converter as determined by the main-beam clutter is

$$N_{\text{AD,amp}} \geq \text{CEIL}\left[\frac{[(C/N)_{\max}]_{\text{dB}} + [\text{fluct\_margin}]_{\text{dB}} + 20\log_{10}\left[[\text{noise}]_{\text{quanta}}\right]}{6}\right] \tag{4.7}$$

where CEIL(x) is the smallest integer $\geq x$. The instantaneous dynamic range supported by an A/D converter improves about 6 dB per bit.[47]

For the example cited in Figure 4.15, where the maximum C/N is 53 dB at a 1000-ft radar altitude and with a fluctuation margin of 10 dB and thermal noise at 1.414 quanta (3 dB), the A/D converter requires at least 11 amplitude bits plus a sign bit for a total of 12 bits to achieve the peak A/D dynamic range of 63 dB. The upper portion of Figure 4.16 illustrates this case. The lower portion of Figure 4.16 will be used in the stability discussion to follow.

Stability. To achieve the theoretical clutter rejection and target detection and tracking performance of a pulse doppler system, the reference frequencies, timing signals, and signal processing circuitry must be extremely stable.[48–52] In most cases, the major concern is with short-term rather than long-term stability. Long-term stability mainly affects velocity or range accuracy or spurious signals (due to PRF harmonics) but is relatively easy to make adequate. Short-term stability refers to variations within the round-trip radar echo time or during the signal coherent integration time. The most severe stability requirements relate to the generation of spurious modulation sidebands on the main-beam clutter, which raise the system noise floor or can appear as targets at the detectors. Thus, the maximum ratio of main-beam clutter to system noise measured at the receiver output (C/N), including the fluctuation margin as discussed above, is the predominant parameter that determines stability requirements.

Target returns compete with clutter returns and noise for detection. Suppose desired targets have sufficient radial speed so that they lie in the clutter-free region of doppler frequency when a pulse doppler waveform is used. These targets now have to compete only with system noise. This noise can be both additive and multiplicative. Additive noise tends to mask multiplicative noise in low-performance radars.

Additive noise sources can be external to the radar, such as atmospheric noise (sky temperature), ground noise (black body radiation), and jammers, or they can be internal, such as thermal noise. Thermal noise is also known as *Johnson noise* and

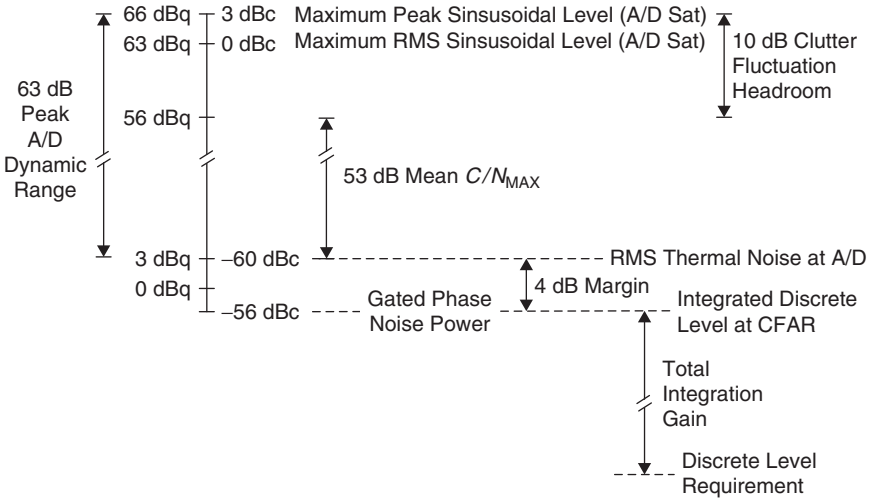


FIGURE 4.16 Dynamic range and stability levels

gaussian noise, the latter term arising from the gaussian statistics of its voltage probability density function. Thermal noise is always present in the radar receiver and is the ultimate limit on radar sensitivity. The absolute level of additive noise sources is determined by the source and its relation to the radar. Proper system design can reduce thermal noise to a level where multiplicative noise can become significant in limiting the radar sensitivity.

Multiplicative noise is characterized by either a time-varying amplitude (amplitude modulation, AM) or a time-varying phase (phase modulation, PM, or frequency modulation, FM). The absolute level depends on the strength of the signal (carrier) on which the noise source is riding. Multiplicative noise sources are frequency instabilities, power supply ripple and noise, 1/*f* noise, timing jitter, and unwanted mixer products (*discretes* or *spurs*). Multiplicative noise modulates radar returns by varying their amplitude or phase and is present on all radar returns being most apparent on large returns such as main-beam clutter. The result in the spectral domain is spurious modulation sidebands. Random multiplicative noise broadens the spectrum of the carrier frequency. Discrete multiplicative noise sources generate discrete spectral lines that can cause false alarms.

System stability is characterized by the overall two-way (transmit and receive) composite system frequency response, which is the return of a nonfluctuating target as a function of doppler frequency. System frequency response should be defined by the doppler passband.[53] The focus of this section will be the stability requirements for doppler frequencies separated enough from the carrier to be outside the ground moving target notch. The concern in this region is white phase noise, which determines the phase noise floor. Low frequency (i.e., closer to the carrier) stability is more applicable to air-to-ground pulse doppler modes such as GMTI and SAR.

The location of an instability source within the system will determine whether it is imparted upon a return signal via the transmit path, receive path, or both. Instabilities either on transmit or receive are called *independent*. Those imposed on both transmit and receive are *common*.

Amplitude instabilities caused by AM tend to be considered independent since the LOs drive the mixers in the receiver into compression. Also, transmitters work most efficiently when driven into compression (i.e., where the power amplifier is saturated and provides a constant output power level regardless of small deviations on the input). Instabilities due to PM (of which FM is a special case) tend to dominate those due to AM. As such, the focus will be on phase disturbances: random phase noise and discrete sinusoidal signals (spurious signals).

Random Phase Noise. Random phase noise riding on a large signal can mask weak target returns. The object is to specify system phase noise so that it is well below the thermal noise when a large signal at the A/D saturation level is present in the receiver. (A signal at A/D saturation is the largest signal that can be linearly processed by the radar receiver.) Then the radar sensitivity is limited by thermal noise (always present) plus a small increase in the total noise level caused by the phase noise.

The phase noise of oscillators and other components is typically specified as the multiplicative noise that rides on a continuous waveform, or CW phase noise. In pulse doppler radar, transmit gating interrupts the continuous waveform to produce a pulsed waveform. *Gated* phase noise is the result of gating CW phase noise. The spectrum of a pulsed (gated) signal is different from CW. The resulting noise, the gated noise, can be much different from the CW noise, especially for low duty cycle waveforms and noise close to the carrier. It is preferable to make noise measurements on equipment

under the same gating conditions that will be used in the radar system. Some devices, such as high power transmitters, cannot operate continuously and only gated noise measurements are possible. The gated phase noise spectrum is the summation of the CW phase noise spectrum replicas centered at frequencies $\pm nf_R$, where f_R is the PRF and n is an integer. The total gated phase noise in the PRF bandwidth f_R equals the total CW phase noise in the transmit pulse bandwidth. In terms of stability requirements, the system requirements are derived using gated phase noise, which in turn is converted to a CW value for specifying components such as oscillators. The CW phase noise floor is smaller by a factor of the ratio of the PRF to the transmit bandwidth when the CW phase noise is assumed to be white.

Sensitivity loss due to phase noise is quantified by the increase in the system noise floor in the "clutter-free" doppler filters due to the phase noise sidebands on a large signal such as main-beam clutter. Sensitivity loss is the amount by which the total noise (thermal plus phase) exceeds the thermal noise level, as shown in Eq. 4.8. A gated phase noise to thermal noise ratio of −4 dB results in an approximately 1.5 dB sensitivity loss. This assumes a worst-case scenario with the main-beam clutter return at the A/D saturation level. CAGC, discussed in Section 4.1, is typically used to regulate the mean clutter to a level below A/D saturation (typically by the amount of the expected clutter fluctuation level). With CAGC, sensitivity loss will be less than or equal to the calculated worst-case value.

$$[\text{Sensitivity Loss}]_{\text{dB}} = 10\log_{10}\left(1 + \frac{\text{Gated Phase Noise Power Density}}{\text{Thermal Noise Power Density}}\right) \tag{4.8}$$

Table 4.5 contains a calculation of the phase noise floor requirements for an 180 kHz PRF waveform. Clutter levels that require a 12-bit (sign plus 11 amplitude bits) A/D converter are assumed, as shown in Figure 4.16. The transmit pulse duration is 1.75 μs, resulting in a transmit pulse bandwidth of approximately 0.5 MHz since no pulse compression is used. The rms thermal-noise power is the thermal-noise floor within the receive portion of IPP. This power level is given in decibels with respect to the carrier amplitude (dBc). The thermal-noise density is obtained by dividing this power by the PRF bandwidth. The maximum gated phase noise floor is set to be 4 dB below the thermal noise floor for at most a 1.5 dB sensitivity loss. The CW phase noise floor is then obtained by multiplying by the PRF to transmit bandwidth ratio.

TABLE 4.5 CW Phase Noise Density Floor Calculation

Parameter	Value [dB]	Units	Comment
Thermal Noise Power at A/D	−60.0	dBc	12-bit A/D (sign+11 bits) thermal noise set at 1.414 quanta
1/PRF Bandwidth	−52.6	dB/Hz	180 kHz PRF waveform
Thermal Noise Density Floor at A/D	**−112.6**	**dBc/Hz**	
Phase Noise to Thermal Noise Ratio	−4.0	dB	Margin for at most 1.5 dB sensitivity loss
Gated Phase Noise Density Floor	**−116.6**	**dBc/Hz**	
PRF-to-Transmit Bandwidth Ratio	−5.0	dB	0.5 MHz transmit pulse bandwidth 1.75 μs pulse width w/ no PC
CW Phase Noise Density Floor	**−121.6**	**dBc/Hz**	

TABLE 4.6 Notional Subsystem Phase Noise Allocation

Subsystem		Allocation		Adjustment for Common Source [dB]	Requirement [dBc/Hz]
		Percentage	dB		
Transmitter		20.0%	−7.0	0.0	−128.6
Exciter	AM	12.5%	−9.0	0.0	−130.6
	PM	37.5%	−4.3	−3.0	−128.9
Reciver		20.0%	−7.0	0.0	−128.6
Synchronizer		10.0%	−10.0	−3.0	−134.6
System		**100.0%**			**−121.6**

The system-level CW phase noise floor requirement (−121.6 dBc/Hz) is allocated to the contributing hardware units. The percentages are based on experience and negotiations with the subsystem designers. A possible allocation is provided in Table 4.6.

Discretes. Some sources of discrete sidebands are ripple on power supplies and the pickup of digital clocks. It is desirable to keep the integrated discrete sidebands below noise at the CFAR input to prevent detecting these discretes and producing false alarms. All coherent and postdetection integration must be accounted for when we specify discrete phase noise requirements.

Common discretes are affected by the time-delay between the portion imparted on the transmit and that on receive. The time-delay changes the correlation between the phase of the spurious modulating frequency from the transmit path with the phase from the receive path.[54] This can relieve the common discrete level requirement for low-PRF (or MTI) waveforms that are range unambiguous. However, for highly range ambiguous medium-PRF and high-PRF waveforms, the assumption is made that the noise common to transmit and receive adds noncoherently in the downconversion process. As a result, the common discrete power increases by 3 dB.

Table 4.7 provides the calculation for the system requirements for independent and common discrete levels. As in Table 4.5, a maximum clutter level requiring a 12-bit A/D is assumed and the rms thermal-noise level at the A/D converter is set to 1.414 quanta. To form the doppler filters, 2048 pulses are coherently integrated.

TABLE 4.7 Discrete Level Requirement Calculation

Parameter		Value [dB]	Units	Comment
Thermal Noise Power at A/D		−60.0	dBc	12-bit A/D (sign+11 bits) thermal noise set at 1.414 quanta
Total Integration Gain	Number of Pulses Coherently Integrated	33.1	dB	2048 IPPs integrated per CPI
	Dopper Filter Weighting	−2.66	dB	90 dB Dolph-Chebyshev weighting loss
	Number of CPIs Noncoherently Integrated	3.82	dB	PDI of 3 CPIs per Look 10 $\log_{10}(N_{pdi}{}^{0.8})$
Thermal Noise Power at CFAR		−94.3	dBc	Effective noise level after integration
Discrete to Thermal Noise Margin		−4.0	dB	Provides low P_{FA} due to discretes
Independent Discrete Requirement		**−98.3**	**dBc**	
Common Discrete Requirement		**−101.3**	**dBc**	3 dB less than Independent Discrete

To reduce doppler filter sidelobes, a 90 dB Dolph-Chebyshev weighting is applied, which reduces the coherent integration SNR gain by about 2.66 dB. For detection, three CPIs are integrated noncoherently via PDI for an approximate integration gain in dB of $10 \log_{10}(N_{\text{PDI}}{}^{0.8})$, or 3.82 dB. This results in a thermal-noise level of −94.3 dBc at the detector. A discrete to thermal-noise margin of −4 dB is used to provide a low P_{FA} due to discretes. The common discrete requirement is made 3 dB more stringent relative to the independent requirement as discussed above.

4.4 RANGE AND DOPPLER AMBIGUITY RESOLUTON

Medium and high-medium PRF waveforms usually use multiple discrete PRF ranging to resolve range ambiguities, while linear FM ranging is commonly employed when high-PRF waveforms are used.

Multiple Discrete PRF Ranging. The techniques for calculating true range from several ambiguous measurements generally involve sequential measurement of the ambiguous range in each PRF, followed by an *unfolding* and correlation process. The unfolding creates a vector of possible ranges for each valid detection by adding a set of integers [0 … K] times the unambiguous range interval:

$$R_{\text{unfold}} = R_{\text{ambiguous}} + \frac{c}{2f_R}[0 \ldots K] \tag{4.9}$$

where the unambiguous range interval = $c/(2f_R)$, with c = speed of light and f_R = PRF. The set of integers [0…K] are referred to as the range ambiguity numbers, with K determined by the maximum range of interest (K = CEIL[$2R_{\max} f_R/c$]). Range correlation occurs when the unfolded detections are scanned and a correlation window is applied across looks, as shown in Figure 4.17. In this example, the correlated target range has an ambiguity number of 4 (5th time around echo) on PRF 1, and an ambiguity

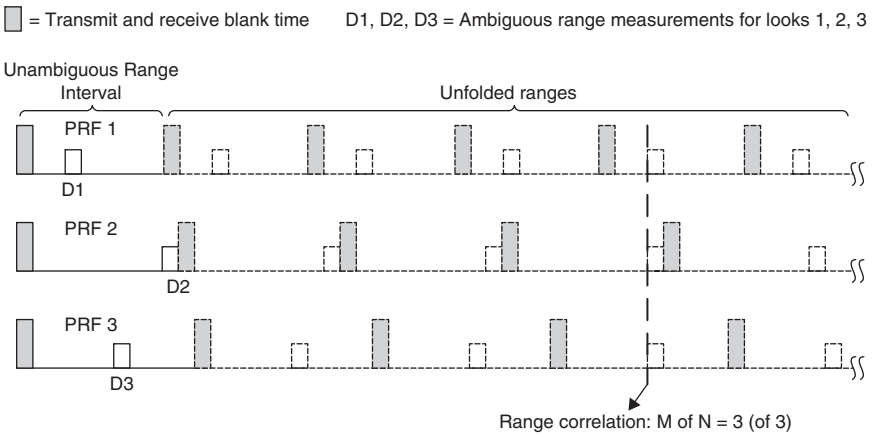


FIGURE 4.17 Range correlation example with 3 PRFs

number of 3 on PRFs 2 and 3. The IPP lengths (often expressed in *range gates per IPP*) are usually kept relatively prime (no common factors except the number 1) to permit unambiguous ranging at the maximum possible range.

The logic for correlation requires at least M detections across the N PRFs in a dwell to declare a target report (with M typically ≥ 3 for medium- and high-medium PRF waveforms). Range ghosts occur if the correlated range does not represent the true target range and typically occur when there is more than one detection per look. Range ghosts can also occur if a target detection on a single look correlated with other dissimilar target(s), or if multiple range correlations occurred on a set of detections corresponding to a single unique target (i.e., multiple unfolded ranges fell within the correlation window).

One method for efficiently scanning and correlating the unfolded detections involves *coarse binning,* as shown in Figure 4.18. Here, ambiguous detections are first amplitude centroided and then unfolded, as discussed previously, but with the results stored in an array whose elements are the coarse bins. These bins have a size less than or equal to the shortest IPP, and correlation involves scanning identical bins across all of the PRFs in the dwell and applying a correlation window. In the example shown in Figure 4.18, the bins are set to nine range gates (shortest IPP length), and the fifth coarse bin contains detections across the three PRFs that fall within the correlation window of ± 0.3 range gates. Blank, or empty, bins occur when the unfolded range falls outside a particular coarse bin interval. Key advantages to this approach are the ability to change the range correlation window dynamically and perform motion compensation easily for the range change across the dwell due to radar platform motion and/or the target's motion (if the unambiguous doppler has been resolved prior to this process). Additionally, the range gate sizes do not need to stay the same across the set of PRFs used in the dwell; in this case, the ambiguous range gate measurements on each look are first converted to common distance units (e.g., meters) prior to the unfolding and scanning/correlation processes.

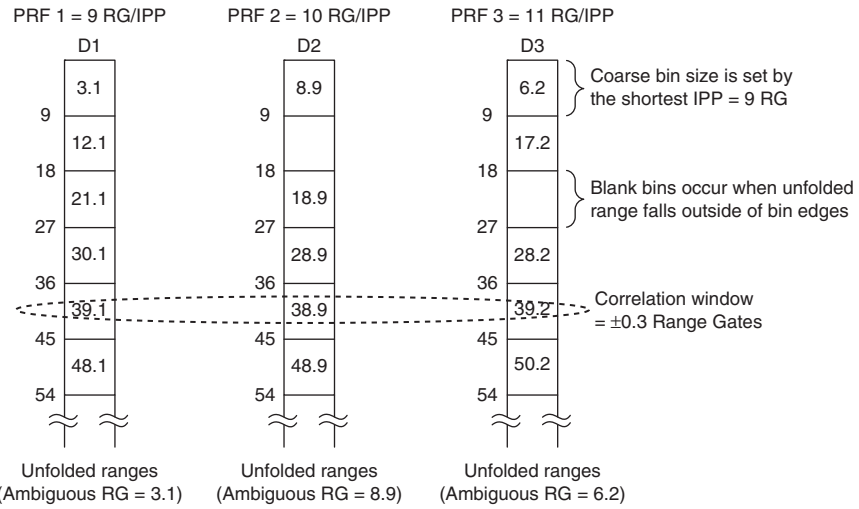


FIGURE 4.18 Range correlation using coarse binning on unfolded, centroided ambiguous detections. In this example, range gate size is the same for all three PRFs.

Additional criteria can be used to reject range ghosts, such as selecting the correlated range with the highest *M-of-N* value, selecting the detections based on the minimum variance across the *M* detections, or using maximum likelihood techniques.[55] The computed radar cross section (RCS) of correlations can also be used in the correlation process to reject sidelobe discrete detections as described in Section 4.2 (postdetection STC).

The ghosting problem can be mitigated further by a combination of doppler and/or monopulse binning. Resolving the doppler ambiguities first (prior to range correlation) will reduce the set of detections to those within the doppler correlation window. For cases where this is not feasible (generally the lower medium PRFs), utilizing both range and doppler correlation will reduce ghosts. Using monopulse measurements to segregate and bin targets that are distinguishable in angle can also reduce ghosting when there are a significant number of detections in a dwell.

A typical medium- or high-medium-PRF pulse doppler waveform cycles through *N* unique PRFs in a processing dwell (*N* typically being 5 to 8). The medium PRFs generally cover nearly an octave in frequency for good doppler visibility and ground moving target rejection. However, high-medium PRFs have inherently good doppler visibility (since they are ambiguous in sign only), so the span of the PRFs in a set of *N* PRFs is usually much less than an octave. Additional constraints on PRF selection for both waveforms include good visibility in sidelobe clutter (where some PRFs may be obscured by clutter in portions of the ambiguous range interval) and minimization of ghosts in the ambiguity resolution processing.

Doppler Ambiguity Resolution. Resolution of the unambiguous doppler-velocity is needed for medium-PRF waveforms, and it is generally done with a similar unfolding and correlation technique, as described previously for range ambiguities. As shown in Figure 4.19, velocity unfolding of detections involves adding a set of signed integers

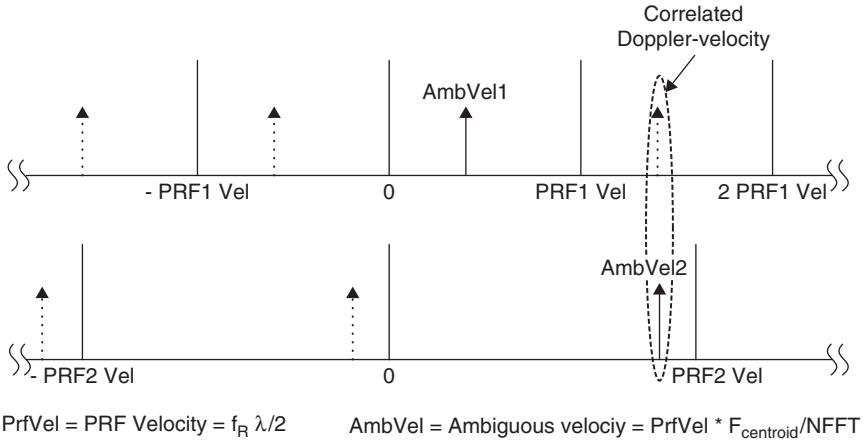


FIGURE 4.19 Doppler-velocity correlation performed on two detections across two looks. Ambiguous detections are unfolded out to a maximum positive and negative velocity.

times the PRF *velocity* (first blind speed) to each measured ambiguous radial velocity as follows:

$$V_{\text{unfold}} = \frac{f_R \lambda}{2}\left(\frac{F_{\text{centroid}}}{N_{\text{FFT}}} + [-J \ldots 0 \ldots K]\right) \tag{4.10}$$

where $f_R\lambda/2$ is the first blind speed (PRF velocity), F_{centroid} is the amplitude-centroided doppler filter number, N_{FFT} is the number of filters in the doppler filter bank, and $[-J \ldots 0 \ldots K]$ represents the set of doppler ambiguity numbers covering the maximum negative and positive doppler-velocities for the targets of interest. For cases where there are only a few ambiguities in doppler, doppler correlation may be performed prior to or in conjunction with range correlation to minimize ghosting.

High-PRF Ranging. Range-ambiguity resolution in high PRF is performed by modulating the transmitted signal and observing the phase shift of the modulation on the return echo. Modulation methods include varying the PRF, either continuously or in discrete steps; varying the RF carrier, with either linear or sinusoidal FM; or some form of pulse modulation such as pulse-width modulation (PWM), pulse-position modulation (PPM), or pulse-amplitude modulation (PAM). Of these modulation techniques, PWM and PPM may have large errors because of clipping of the received modulation by eclipsing or straddling (discussed in Section 4.6), and PAM is difficult to mechanize in both the transmitter and the receiver. Consequently, they will not be further considered here.

Linear-Carrier FM. Linear frequency modulation (FM) of the carrier can be used to measure range. The modulation and demodulation to obtain range are the same as used in frequency-modulated continuous-wave (FM-CW) radar,[56] but the transmission remains pulsed.

Suppose the dwell time is divided into two looks. In the first look, no FM is applied, and the doppler shift of the target is measured. In the second look, the transmitter frequency is varied linearly at a rate $\dot{f}$ in one direction (i.e., increasing or decreasing in frequency). During the roundtrip time to the target, the local oscillator has changed frequency so the target return has a frequency shift, in addition to the doppler shift, that is proportional to range. The difference in the frequency Δf of the target return between the two looks is found, and the target range calculated from

$$R = \left|\frac{c\Delta f}{2\dot{f}}\right| \tag{4.11}$$

The problem with only two FM segments during a dwell is that, with more than a single target in the antenna beamwidth, range ghosts result. For example, with two targets present at different dopplers, the two frequencies observed during the FM period cannot be unambiguously paired with the two frequencies observed during the no-FM period. To mitigate this problem, a three-segment scheme is used with the following segments: no-FM, FM-up, and FM-down. The range is found by selecting returns from each of the three segments that satisfy the relations

$$f_1 < f_0 < f_2 \tag{4.12}$$

$$f_1 + f_2 = 2f_0 \tag{4.13}$$

TABLE 4.8 Three-slope FM Ranging Example

There are two targets, *A* and *B;* FM slope = 24.28 MHz/s.

Target	***A***	***B***
Range (nmi)	10	20
Doppler frequency (kHz)	21	29
FM shift (kHz)	3	6
Observed Frequencies		
f_0, no FM (kHz)	21	29
f_1, FM up (kHz)	18	23
f_2, FM down (kHz)	24	35

Possible sets that satisfy the relations shown in Eq. 4.12 and Eq. 4.13 are

f_1	f_0	f_2	$2f_0$	$f_1 + f_2$	**Target?**	**Range (nmi)**
18	21	24	42	42	Yes	10
18	21	35	42	53	No	
18	29	35	58	53	No	
23	29	35	58	58	Yes	20

where f_0, f_1, and f_2 are the frequencies observed during the no-FM, FM-up, and FM-down segments, respectively. The range then is found from Eq. 4.11, where

$$\Delta f = f_2 - f_0 \quad \text{or} \quad (f_2 - f_1)/2 \quad \text{or} \quad f_0 - f_1 \tag{4.14}$$

An example is shown in Table 4.8.

If more than two targets are encountered during a dwell time, ghosts again result, as only $N - 1$ simultaneously detected targets can be resolved ghost-free where N is the number of FM slopes. However, this problem is not severe in practice, since multiple targets in a single beamwidth are usually a transient phenomenon.

The accuracy of the range measurement improves as the FM slope increases since the observed frequency differences can be more accurately measured. The FM slope is, however, limited by clutter-spreading considerations, since during the FM periods, the clutter is smeared in frequency and can appear in frequency regions normally clear of clutter.[57] A no-FM, FM-up, double FM-up scheme is recommended to prevent desired targets from competing with main-beam clutter. Range accuracies on the order of 1 or 2 miles can be reasonably achieved.

4.5 MODE AND WAVEFORM DESIGN

Modern multifunction pulse doppler radars utilize various modes to accomplish tasks such as search and track. Each mode uses certain waveforms optimized for the detection and measurement of various target characteristics.

For example, the radar operator might select a search mode and specify a search volume that the radar will raster scan, as shown in Figure 4.7. Valid detections in search are then converted to tracks in the radar computer. These tracks need to be updated by a track mode on a regular basis depending on the track accuracy required. High track accuracy is needed for threatening targets or those that need a fire control

solution in order to engage, as opposed to nonthreatening targets where a general situational awareness is sufficient and high accuracy is not required.

Search. The two primary search modes are *Autonomous Search* and *Cued Search*. In Autonomous Search the operator selects a range, azimuth, and elevation coverage, and the radar searches each beam position that covers this volume once per frame. The time it takes to complete a frame is known as the revisit or frame time. The *frame time* should be minimized to enhance the cumulative probability of detection of targets.

Modern radar systems can take advantage of on- and off-board cues to increase the probability of acquiring a target using Cued Search. A Cued Search mode adjusts the search volume and waveform selection according to the accuracy of the cue's parameters.

Radars with electronically scanned array (ESA) antennas can interleave other functions (track updates, Cued Search, calibrations, etc.) with Autonomous Search. The radar computer's resource manager must ensure that the maximum frame time is not exceeded with the inclusion of these other functions during a search frame.

For airborne pulse doppler radars, Autonomous Search can have two submodes: *Forward-aspect* and *All-aspect Search*. Forward-aspect Search is designed to detect head-on engagement targets with high closing speeds that are not competing against main-beam or sidelobe clutter. Forward-aspect Search uses high-duty high-PRF waveforms to maximize the energy on target and provide long detection range. Forward-aspect Search waveforms include *Velocity Search* (VS), *High-PRF Range-While-Search* (HRWS), and *Alert/Confirm*. All-aspect Search can be either a single high-medium PRF waveform that has acceptable performance for targets that are competing with sidelobe clutter, or the combination of Forward-aspect Search high-PRF waveforms interleaved with medium-PRF waveforms designed to detect targets competing with sidelobe clutter, such as *Medium-PRF Range While Search* (MRWS).

Velocity Search. VS is a high-PRF search waveform that measures doppler frequency unambiguously (with the possible exception of sense), but does not measure range. This is the classic high-PRF waveform. The transmit duty cycle is maximized to increase detection range. The receiver may be range gated to match the bandwidth of the transmit waveform, but range measurement is not attempted.

A VS dwell will consist of a single look at a given PRF. The coherent integration time is maximized within the limits of the maximum expected target radial acceleration. VS is optimized for Swerling I and III target amplitude fluctuation statistics and the cumulative probability of detection of incoming targets over several search frames.

High-PRF Range-While-Search. Like VS, HRWS is a high-PRF waveform. However, linear-carrier FM ranging is used to obtain a range measurement, as described in Section 4.4. This range measurement comes at the expense of frame time with the addition of various FM slopes for each dwell. The accuracy of this range measurement is dependent upon the linear FM ranging slopes.

Alert/Confirm. The beam agility of ESA-based radars allows the use of sequential detection techniques.[58] A simplification of such techniques is known as Alert/Confirm.[59,60] The goal of Alert/Confirm is to provide high sensitivity while managing false alarms and minimizing the search frame time. By transmitting a longer *Confirm* dwell for ranging only at beam positions where a shorter-dwell *Alert* has detected

targets, Alert/Confirm provides the range measurement of classic HRWS waveforms without the frame time expense of transmitting linear FM ranging dwells every beam position. The Confirm dwell can also be used to control false alarms, permitting the Alert dwell to be more sensitive than classic VS.

The Alert phase is used to search each beam position of the frame for the presence of a target. A VS waveform is used with a low detection threshold and a corresponding false alarm time on the order of a few seconds. The lower detection threshold increases sensitivity. When an Alert dwell declares a detection, a Confirm dwell is scheduled for that Alert dwell's beam position. If monopulse measurements are available on the Alert detection, the Confirm beam can be centered on the detection to decrease beam-shape loss. The Confirm dwell is typically a HRWS waveform and only examines doppler filters within a window centered about the filter of the Alert detection cue. The Confirm dwell must produce a detection corresponding to the Alert detection in order for a valid detection declaration. The Confirm dwell is used to manage false alerts and provide a range measurement for target detections. The Alert and Confirm detection thresholds are designed to achieve overall false alarm time equal to conventional search (one every few minutes). Along with using the same PRF in Alert and Confirm, the time between these dwells, or *latency*, should be minimized to prevent a valid Alert detection from being eclipsed during the Confirmation dwell.

Low-latency also allows the use of *Correlated* Alert/Confirm. Here, a Swerling I target RCS fluctuation model is assumed. This implies that when the same RF carrier frequency is used for Alert and Confirm, the target RCS will be relatively constant between the two dwells,[61] providing additional range enhancement in terms of the cumulative probability of detection.

Medium-PRF Range While Search. A medium-PRF waveform is used to detect targets competing with sidelobe clutter that would be undetectable in HRWS. MRWS allows the detection of nose aspect targets at wide scan angles that are crossing the radar line-of-sight, such that their low closing velocity places them in sidelobe clutter and tail aspect targets in lead pursuit engagements (an attack geometry where the nose of the attacking aircraft is pointed ahead of the target's present position). MRWS provides complete situational awareness (perception of the surrounding tactical environment), but does not have the maximum detection range provided by the higher duty cycle of HRWS for thermal noise-limited targets.

The MRWS waveform uses *M*-of-*N* detection processing; a typical waveform might be 3-of-7. Each MRWS dwell is made up of *N* looks each with a different PRF. Detection is required on at least *M* looks to resolve target range and range rate unambiguously. The detection thresholds are set to provide approximately one false alarm per minute.

The effectiveness of MRWS is dependent on the ability to detect targets at the required ranges while simultaneously rejecting discrete clutter detections. Low two-way antenna sidelobes along with the combination of techniques discussed in Section 4.2, such as guard channel blanking and postdetection STC, are used to mitigate sidelobe clutter discrete false alarms.

MRWS also uses pulse compression to decrease the amount of sidelobe clutter that targets must compete with. The lower PRF reduces eclipsing and the amount of clutter range-folding. Transmit carrier frequency diversity dwell-to-dwell forces Swerling I and III target fluctuation statistics and improves cumulative probability of detection performance. Frequency diversity look-to-look within a dwell produces Swerling II and IV statistics and is better suited for high single-scan probability of detection.

MRWS can also be implemented with a high-medium PRF, which is characterized by the waveform's doppler coverage being unambiguous in doppler magnitude, but not doppler sense, for the maximum target doppler of interest. The resulting single blind speed due to main-beam clutter permits as wide of a clutter rejection notch as required to reject main-beam clutter or ground moving targets and still not result in doppler blind speeds for targets of interest. *M*-of-*N* ranging provides better range measurement accuracy than linear FM ranging used in HRWS. The PRFs used in a dwell must be chosen to resolve the high number of range ambiguities within the instrumented range.

Track. Target tracking is done by making range, range rate, and azimuth and elevation angle measurements on targets. Range measurements are obtained using range gating and centroiding on the target return with range ambiguities being resolved within the tracker. Range rate (i.e., doppler) measurements are formed with a centroid on the target's doppler return in the filter bank. Angle measurements can be obtained using monopulse, sequential lobing, or conical scan, with monopulse being the prominent choice in modern radars. The tracker creates windows, or groups of contiguous range-doppler cells, around each of these measurements in order to associate detections with existing tracks. The tracker, usually implemented with a nine-state (position, velocity, and acceleration) Kalman filter, estimates target motion in an inertial coordinate system.

Multiple-Target Tracking (MTT) can be accomplished in several ways. One method (*Track-While-Scan,* or TWS) is to use the normal search mode with FM or multiple-PRF ranging and store the range, angle, and doppler of the reported detections in the radar computer. These detections are then used to form and update track files. The antenna scans in a normal search pattern, and a scan-to-scan correlation is made on the detections that update the track files. Although tracking accuracies are less than can be achieved in a dedicated single-target track, multiple targets can be tracked simultaneously over a large volume of space.

A second method of Multiple-Target Tracking, *Pause-While-Scan,* particularly applicable to ESA-based radars, is to scan in a normal search pattern, pause on each search detection, and enter a Single-Target Track mode for a brief period. The advantage is that the resulting range, angle, and doppler measurements are more accurate than those made with a scanning antenna, but the time to search a volume in space is increased.

Transition-to-Track, or Track Acquisition, is used to confirm search target detections and provide improved range accuracy when needed. If the target is successfully acquired, a track file in the radar computer is initiated. The Track Acquisition waveform's parameters depend upon the type of search waveform that produced the target detection. The Track Acquisition waveform's thresholds are set to reject false alarms and reduce the false track initiation rate to less than one per hour.

For Track Acquisition, a search detection from VS would require a HRWS waveform to obtain a range measurement. HRWS and Alert/Confirm waveforms are followed by range gated high-PRF dwells using *M*-on-*N* ranging to achieve the necessary range accuracy for single PRF track updates. The unambiguous HRWS range measurement of the search detection is used to help resolve the range ambiguity. For MRWS detections, another MRWS dwell is used for Track Acquisition. Once the track file is initiated, several rapid track updates are used to firmly establish the track.

When doing Single-Target Track updates, a single PRF waveform can be used. The range and/or doppler ambiguities are resolved in search and, if necessary, in the Transition-to-Track phase. By using the unambiguous range and velocity predictions

of the target provided by the tracker, a single PRF can be chosen such that range and doppler eclipsing is avoided with high probability. The length of the dwell is adapted to provide sufficient energy on target so that its return signal-to-noise ratio will provide the necessary measurement accuracies required by the tracker. This adaptive track update waveform allows the search revisit time to be maintained while tracking multiple targets.

4.6 RANGE PERFORMANCE

The radar range equation is used to determine the performance of pulse doppler radar. The radar range equation must include losses, both system and environmental, that will degrade the strength of return signals at the detector. Probability of detection (P_d) depends on target signal-to-noise ratio and probability of false alarm (P_{FA}), which itself is a function of waveform. The false alarm probability determines the detection threshold and is referenced to an individual range-doppler cell. This per-cell probability is derived from the specified false report time for the system.

Radar Range Equation. In the doppler region where the signal does not fall in clutter, performance is limited only by system noise. The signal-to-noise power ratio in the range-doppler cell at the detector prior to postdetection integration for a target at range R is given by

$$\text{SNR} = \left(\frac{R_o}{R}\right)^4 \tag{4.15}$$

$$R_o = \left(\frac{P_{\text{av}} G_T G_R \lambda^2 \sigma_T}{(4\pi)^3 k T_s B_n L_T}\right)^{1/4} \tag{4.16}$$

where R_o = range at which SNR is equal to 1
σ_T = target radar cross section
L_T = losses applicable to the target

The remaining terms are as defined following Eq. 4.2. The net loss L_T used to compute SNR for a target is generally higher than the net loss L_C used to compute CNR in Eq. 4.2. L_T includes losses, such as eclipsing and range gate straddle, doppler filter straddle, CFAR, and guard blanking, that are applicable to resolvable targets but not to distributed clutter.

The target SNR represents the envelope ($\sqrt{I^2 + Q^2}$ for a linear detector or $I^2 + Q^2$ for a square-law detector) of the target return compared to that of just noise. The envelope is measured after the entire coherent matched filter process (i.e., transmit pulse matched filter, pulse compression, and coherent doppler filtering). Therefore, SNR is associated with a single CPI.

Losses. Some of the losses inherent in, but not necessarily unique to, pulse doppler radars that employ digital signal processing are discussed below. Some of the losses may be incorporated into the other variables in the radar range equation. Care must be taken to account for all of the system losses while avoiding redundancies.

Most *front-end* losses are applicable to both targets and clutter. Losses applicable only to targets will be indicated.

RF Transmit Loss. This loss accounts for RF ohmic losses between the transmitter or RF power amplifier and the antenna radiator, which can include losses from connectors, circulators, and radiating elements.

Radome Loss. Most radars require a radome to protect the antenna from environmental elements and to conform to the platform's shape. Radomes will have a loss that may depend on the scan angle of the antenna. This loss must be accounted for on transmit and receive (i.e., a two-way loss).

Propagation Loss. Propagation through the atmosphere results in a loss, especially at higher radar carrier frequencies. This loss is a function of range, altitude, and weather. It is also a two-way loss. Propagation loss is more of a environmental loss than a system loss, but can be grouped with the other losses that make up net loss in the radar range equation.

Scan Loss. Broadside electronically scanned array antennas are subject to reduction in gain when the main beam is scanned off broadside. The projected area of the ESA aperture decreases as beam scans from broadside. Projected area drops as cosine of scan (cone) angle. Mutual coupling between radiating elements further reduces the effective area. Scan loss must be accounted for on transmit and receive.

Beamshape Loss. This target-specific loss accounts for the loss in gain when the target is not located at the peak of the beam. Beamshape loss is defined as the increase in the power or the SNR required to achieve the same probability of detection on a target spread uniformly over the specified beam coverage as would occur with a target at beam center. Beamshape loss is used primarily in search detection range performance calculations.

RF Receive Loss. This loss is similar to RF Transmit Loss except it accounts for ohmic losses from the antenna face to the first low-noise amplifier. This loss may be included in the receive system noise figure or system temperature value.

IF Matched Filter Loss. The matched filter for a pulse doppler waveform includes the analog IF matched filter in the receiver and any subsequent digital integration of A/D samples to match the duration of the transmit pulse. IF matched filter loss quantifies how well the analog IF matched filter compares to the ideal matched filter for that point in the reception chain.

Quantization Noise Loss. This loss is due to the noise added by the A/D conversion process and to truncation due to finite word lengths in the signal-processor that follow.[62] This loss can also be incorporated into the receiver noise figure value.

Pulse Compression Mismatch Loss. This is caused by the intentional mismatching of the pulse compression filter to reduce time (range) sidelobes.

Eclipsing and Range Gate Straddle Loss. The large amount of range ambiguity inherent in pulse doppler waveforms results in the possible eclipsing of target returns

when the receiver is blanked during pulse transmission. In a multiple range gate system, the returns may also straddle gates reducing the pulse matched filter output of a single gate. Because of eclipsing and range gate straddle, the value of R_o, given by Eq. 4.16, may fall anywhere between zero and a maximum value, depending on the exact location of the target return within the interpulse period.

Figure 4.20 illustrates the effect of eclipsing and range gate straddle on the output of the pulse matched filter over the IPP. Each range gate is assumed to be matched to the transmit pulse bandwidth, which for unmodulated pulses (i.e., no pulse compression modulation) is the inverse of the pulse duration. Therefore, referring to Figure 4.6, the gate width τ_g equals the transmitted pulse τ_t. In Figure 4.20, the IPP is $5\tau_g$. The plots on the left represent a range gate spacing of τ_s equal to τ_g. Range gate straddle loss can be reduced by the use of overlapping gates at the expense of extra hardware and processing. The rightmost plots represent the use of 50% range gate overlap ($\tau_s = \tau_g/2$). The maximum pulse matched filter output as a function of return delay is shown in terms of relative voltage and power. The "voltage" plot shows the cumulative effect of convolving the return pulse with the matched filter of each range gate. For a single range gate, this is simply the convolution of two rectangular pulses, which results in a triangular response. To compute loss, the matched filter output in terms of power (i.e., voltage squared) must be used.

When the PRF is high, so that many range ambiguities occur, the target range delay may be considered to be random from frame to frame, with a uniform distribution over the IPP. A measure of performance reduction due to eclipsing and range gate straddle is found by

1. Using the uneclipsed detection curve (P_d vs. S/N) for the waveform and selecting a particular SNR of interest S/N_0 along with its corresponding probability of detection $P_{d,0}$.
2. Reduce S/N_0 by a factor related to the relative output "power" of the matched filter as a function of ambiguous range within the IPP. (See the third row of plots in Figure 4.20.)
3. With the reduced SNR, determine the new P_d as a function of ambiguous range within the IPP from the uneclipsed detection curve.
4. Average these new P_d values across the IPP.

The result will be a new detection curve including the average effect of eclipsing and range gate straddle. For a fixed P_d, the difference in SNR between the uneclipsed and the eclipsed detection curves is the average eclipsing and range gate straddle loss. This difference represents the average increase in signal-to-noise ratio required to obtain the same probability of detection with eclipsing and straddle, as in the case where the transmit pulse is received by a matched gate with no straddle. Since the detection curve changes shape, the loss depends on the probability of detection selected, which is depicted in Figure 4.21. For accurate results, eclipsing and range gate straddle loss must be computed together.

A less accurate approximation compares the average signal-to-noise ratio over the interpulse period with the signal-to-noise ratio of the matched case. In the case of N continuous range gates spanning the duration of the IPP, each of which are matched to the transmit pulse width, the approximate average eclipsing and straddle loss is[63]

$$\text{approximate eclipsing and range gate straddle loss} = \frac{12N}{7N-6} \tag{4.17}$$

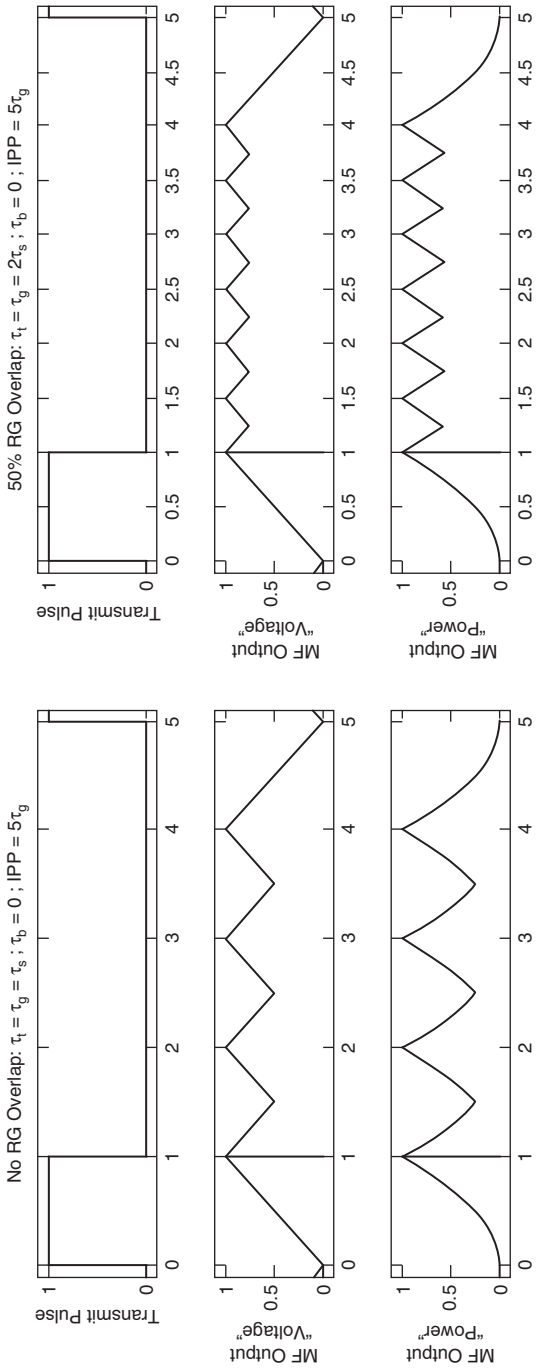

No RG Overlap: $\tau_t = \tau_g = \tau_s$; $\tau_b = 0$; IPP = $5\tau_g$
50% RG Overlap: $\tau_t = \tau_g = 2\tau_s$; $\tau_b = 0$; IPP = $5\tau_g$
Transmit Pulse
MF Output "Voltage"
MF Output "Power"
Time Normalized by Range-Gate Duration

FIGURE 4.20 Concept of eclipsing and range gate straddle loss. The top row of plots shows the transmit pulse for a single IPP of a pulse doppler waveform with a duty cycle of 20%. The second row of plots shows the relative voltage of the maximum pulse matched filter (MF) output as a function of range-ambiguous target return within the IPP. The third row of plots shows the output in terms of relative power.

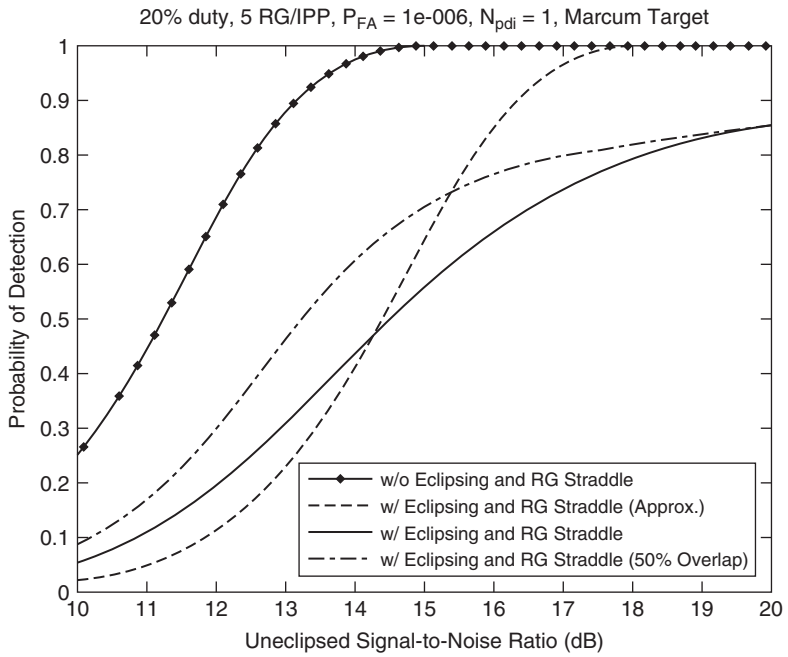

20% duty, 5 RG/IPP, P_FA = 1e-006, N_pdi = 1, Marcum Target
Probability of Detection
Uneclipsed Signal-to-Noise Ratio (dB)
w/o Eclipsing and RG Straddle
w/ Eclipsing and RG Straddle (Approx.)
w/ Eclipsing and RG Straddle
w/ Eclipsing and RG Straddle (50% Overlap)

FIGURE 4.21 Comparison of detection performance with and without eclipsing and range gate straddle loss. The approximate performance using Eq. 4.17 is also provided. The performance with eclipsing and range gate straddle loss with the use of 50% overlapped range gates is shown.

Eq. 4.17 assumes an unmodulated, rectangular transmit pulse shape with the receive gate matched to the transmit pulse width. There is no range gate overlap. The first gate of the N range gates are blanked for the transmit pulse. As shown in Figure 4.21, this approximation is only valid for a P_d near 50%.

There are several other details that have not been included in Figure 4.21. As shown in Figure 4.6, a portion of the first valid receive range gate (and possibly a portion of the last range gate in the IPP) is typically blanked to avoid receiving transients of the transmit-to-receive (and receive-to-transmit) switching. Also, if pulse compression modulation is used on the transmit pulse, the range gate duration will be reduced to match the transmit pulse bandwidth. All of these effects should be included when computing the eclipsing and range gate straddle loss.

Doppler Filter Weighting Loss. This loss results from the increased noise bandwidth of the doppler filters that occurs because of filter sidelobe weighting. The loss can also be accounted for by an increase of the doppler filter noise bandwidth instead of as a separate loss.

Doppler Filter Straddle Loss. This loss is due to a target not always being in the center of a doppler filter. It is computed by assuming a uniformly distributed target doppler over one filter spacing and is a function of the doppler filter sidelobe weighting. This loss can be reduced (at the expense of increased processing) by zero-padding the collected data and performing a higher-point FFT to form highly overlapped doppler filters.

CFAR Loss. This loss is caused by an imperfect estimate of the detection threshold compared with the ideal threshold. The fluctuation in the estimate necessitates that the mean threshold be set higher than the ideal, hence a loss. It is only applicable to targets.

Guard Blanking Loss. This target-specific loss is the detectability loss in the main channel caused by spurious blanking from the guard channel. (See Figure 4.11.)

Probability of False Alarm. Radar detection performance is determined by the detection threshold, which in turn is set to provide a specified probability of false alarm.[64–68] As described in Section 4.4, pulse doppler radars often employ a multilook detection criterion to resolve range ambiguities. This can be accomplished with linear-FM ranging as in the HRWS waveform or *M*-of-*N* ranging used by MRWS. These ambiguity resolution techniques dictate how the probability of false alarm per range-doppler cell is computed. These calculations assume a noise-limited environment.

For HRWS, different linear-FM slopes are applied to looks 2 through *m* of a *m*-look dwell, where *m* is typically 3. The PRF is high enough for at most only a doppler sign ambiguity. Detections in looks 2 through *m*, must correlate in doppler with detections in the first look, which has no slope. A doppler correlation window is set equal to the maximum doppler offset due to linear-FM ranging from a target at the maximum instrumented range. For doppler-only correlation, the P_{FA} per range-doppler cell to provide a specified false report time is

$$P_{FA} = \frac{1}{N_r}\left(\frac{T_d \ln(2)}{\binom{m}{n} N_f N_{FM}^{m-1} T_{FR}}\right)^{1/m} \tag{4.18}$$

where N_r = number of independent range samples processed per IPP
N_f = number of independent doppler filters visible in the doppler passband (number of unblanked filters/FFT weighting factor)
T_d = total dwell time of the multiple PRFs including postdetection integration (if any), space change, and any dead time
n = number of looks in a dwell time
m = number of detections required for a target report (for a typical HRWS dwell, $n = 3$ and $m = 3$)
$\binom{m}{n}$ = binomial coefficient = $n!/[m!(n\text{-}m)!]$
T_{FR} = false-report time (per Marcum's definition where the probability is 0.5 that at least one false report will occur in the false-report time; this can be related to the average time T_{AVG} between false reports by $T_{FR} \approx T_{AVG} \ln(2)$)
$N_{FM} = k_{FM,max}(2R_{max}/c)$ = number of independent doppler filters in the doppler correlation window
$k_{FM,max}$ = steepest linear-FM slope magnitude
R_{max} = maximum instrumented range

Alert/Confirm increases sensitivity by allowing more false alarms in Alert and relying on Confirm to reject those false alerts. The Alert/Confirm combination is designed to provide the same false report time T_{FR} as a conventional waveform. A specified fractional increase F in frame time accounts for the execution of Confirm dwells to reject false Alert detections. F is on the order of 5–10%. When using a VS Alert and a 3-look HRWS Confirm, the probability of false alarm per range-doppler cell, $P_{\mathrm{FA},a}$ and $P_{\mathrm{FA},c}$ for Alert and Confirm, respectively, is

$$P_{\mathrm{FA},a} = \frac{T_{d,a}\ln(2)}{N_{r,a}N_{f,a}T_{\mathrm{FR},a}}$$
$$P_{\mathrm{FA},c} = \frac{1}{N_{r,c}}\left(\frac{2T_{d,c}\ln(2)}{N_{f,\mathrm{cue}}N_{\mathrm{FM}}^2T_{\mathrm{FR}}}\times\frac{F+1}{F}\right)^{1/3} \tag{4.19}$$

where $T_{d,a}$ = total Alert dwell time
$N_{r,a}$ = number of independent range samples processed per IPP in Alert
$N_{f,a}$ = number of independent doppler filters visible in the Alert doppler passband
$T_{\mathrm{FR},a}$ = $T_{d,c}/F$ = Alert false report time
$T_{d,c}$ = total Confirm dwell time
F = fractional increase in frame time allocated to Confirm (5–10%)
$N_{r,c}$ = number of independent range samples processed per IPP in Confirm
$N_{f,\mathrm{cue}}$ = number of independent doppler filters in the Confirm window centered about the doppler of the Alert detection cue
N_{FM} = number of independent doppler filters in Confirm linear-FM ranging doppler correlation window
T_{FR} = overall Alert/Confirm false report time

The M-of-N ranging used in MRWS requires correlation in range and can be viewed as a binary detector. MRWS is typically a medium-PRF waveform with range and doppler ambiguities. Doppler is used for clutter mitigation in each look, and the doppler ambiguity may not need to be resolved since the tracker can determine range rate from successive dwells. A typical MRWS M-of-N correlation would be three detections out of eight looks (i.e., $m = 3$ and $n = 8$). For range-only correlation, the P_{FA} in each range-doppler cell is given by

$$P_{\mathrm{FA}} = \frac{1}{N_f}\left[\frac{T_d\ln(2)}{\binom{m}{n}N_{\mathrm{ru}}T_{\mathrm{FR}}}\right]^{1/m} \tag{4.20}$$

where N_{ru} = number of independent range samples in the output unambiguous-range interval (display range/range gate size)

For better false alarm rejection, doppler correlation can be used for MRWS. In the case where both range and doppler correlation are used, the required P_{FA} is

$$P_{FA} = \left[\frac{T_d \ln(2)}{\binom{m}{n} N_{fu} N_{ru} T_{FR} W^{m-1}} \right]^{1/m} \tag{4.21}$$

where N_{fu} = number of independent doppler filters in the unambiguous doppler region
W = width (in doppler filters) of the correlation window applied to detections following initial detection

Probability of Detection. Using the P_{FA} per range-doppler cell, the probability of detection (P_d) of a given look can be determined for a given target SNR, the number of CPIs noncoherently integrated (N_{pdi}), and the target RCS fluctuation model assumed.[69] The inverse problem of determining the required SNR for a given P_d can be solved via approximations.[70] Universal detection equations have been published that provide reasonably accurate results and are reproduced here.[71] Again, the assumption that targets are in a gaussian noise-limited environment is made.

For a single look with N_{pdi} CPIs noncoherently integrated and a specified P_{FA} per range-doppler cell, the P_d as a function of SNR for a Marcum (nonfluctuating) target can be approximated as

$$P_d(\text{SNR}, P_{FA}, N_{pdi}) = \frac{1}{2} \operatorname{erfc}\left(\sqrt{-0.8 \ln[4P_{FA}(1-P_{FA})]} + \sqrt{\frac{N_{pdi}}{2} - \frac{1}{2}} - \sqrt{N_{pdi}\,\text{SNR} + \frac{N_{pdi}}{2} - \frac{1}{2}} \right) \tag{4.22}$$

where erfc(·) is the complementary error function. The required SNR as a function of P_d for a Marcum target is approximated as

$$\text{SNR}_{reqd}(P_d, P_{FA}, N_{pdi}) = \frac{\eta^2}{N_{pdi}} + \frac{2\eta}{N_{pdi}} \sqrt{\frac{N_{pdi}}{2} - \frac{1}{4}} \tag{4.23}$$

where

$$\eta = \sqrt{-0.8 \ln[4P_{FA}(1-P_{FA})]} - \text{sign}(0.5 - P_d) \sqrt{-0.8 \ln[4P_d(1-P_d)]}$$

For Swerling fluctuating target models, the P_d and required SNR can be approximated, respectively, as

$$P_d(\text{SNR}, P_{FA}, N_{pdi}, n_e) = K_m \left[\frac{K_m^{-1}(P_{FA}, 2N_{pdi}) - 2(N_{pdi} - n_e)}{\frac{N_{pdi}}{n_e}\text{SNR} + 1}, 2n_e \right] \tag{4.24}$$

$$\text{SNR}(P_d, P_{FA}, N_{pdi}, n_e) = \left[\frac{K_m^{-1}(P_d, 2n_e) - 2(N_{pdi} - n_e)}{K_m^{-1}(P_d, 2n_e)} - 1 \right] \frac{n_e}{N_{pdi}} \tag{4.25}$$

where

$$n_e = \begin{cases} 1, & \text{for Swerling I target (chi-squared distribution with 2 degress of freedom)} \\ N_{\text{pdi}}, & \text{for Swerling II target (chi-squared distribution with } 2N_{\text{pdi}} \text{ degress of freedom)} \\ 2, & \text{for Swerling III target (chi-squared distribution with 4 degress of freedom)} \\ 2N_{\text{pdi}}, & \text{for Swerling IV target (chi-squared distribution with } 4N_{\text{pdi}} \text{ degress of freedom)} \end{cases}$$

$$K_m(x,d) = 1 - P\left(\frac{d}{2}, \frac{x}{2}\right) = \text{chi-squared distribution survival function}^{72}$$

$K_m^{-1}(p,d) =$ inverse chi-squared distribution survival function

$$P(\alpha,x) = \frac{\gamma(\alpha,x)}{\Gamma(\alpha)} = \frac{\int_0^x t^{\alpha-1}e^{-t}dt}{\int_0^\infty t^{\alpha-1}e^{-t}dt} = \text{regularized lower incomplete gamma function}$$

The integral of the chi-squared distribution $K_m(x,d)$ and its inverse $K_m^{-1}(p,d)$ are often included in mathematical computation software packages.[73]

When *M*-of-*N* detection (i.e., binary detection) is used within a dwell, the probability of detection for each look ($P_{d,\text{look}}$) is used to compute the probability of detection for a dwell ($P_{d,\text{dwell}}$). When a dwell requires *m* detections out of *n* looks for a target declaration, the $P_{d,\text{dwell}}$ is

$$P_{d,\text{dwell}} = \sum_{k=m}^{n} \binom{k}{n} P_{d,\text{look}}^k \, (1 - P_{d,\text{look}})^{n-k} \tag{4.26}$$

For Alert/Confirm detection performance, the P_d for the Alert dwell and the P_d for the Confirm dwell are individually computed as a function of SNR. Care must be taken to normalize the SNR to account for differences in doppler filter bandwidth between the Alert and Confirm waveforms. The multiplication of normalized probability of detection curve for the Alert dwell with that of the Confirm dwell results in an estimate of the composite P_d vs. *S*/*N* curve for Alert/Confirm. More accurate results must include the effects of latency between the Alert and Confirm dwells.

Search detection performance is often characterized by the cumulative probability of detection, $P_{d,\text{cum}}$, which is defined as the probability that the radar will detect a closing target at least once by the time the target has closed to a specified range. $P_{d,\text{cum}}$ is only defined for closing targets. The cumulative probability of detection for the *k*th scan, or frame, is

$$\begin{aligned} P_{d,\text{cum}}[k] &= 1 - \prod_{i=1}^{k} [1 - P_{d,\text{ss}}[i]] \\ &= P_{d,\text{cum}}[k-1] + P_{d,\text{ss}}[k]\,(1 - P_{d,\text{cum}}[k]) \end{aligned} \tag{4.27}$$

where $P_{d,\text{ss}}[k]$ is the single-scan probability of detection on the *k*th scan. The accumulation of single-scan probability of detections is started at a range where the target's $P_{d,\text{ss}}$ is approximately 5%. There is an optimal search frame time for cumulative detection performance. A balance must be achieved. A short frame time limits the amount of energy placed on target per dwell and lowers the single-scan P_d. A long frame time allows the target to close in range more between revisits, thus lowering the benefit of the accumulation. Figure 4.22 illustrates the difference between single-scan and cumulative probability of detection.

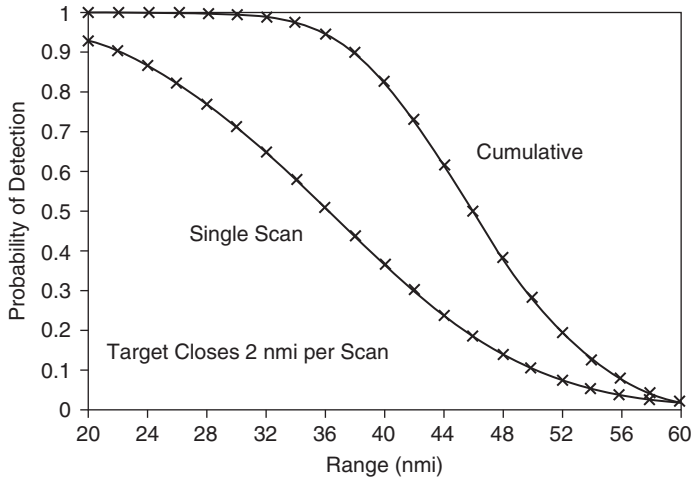


FIGURE 4.22 Single-scan vs. cumulative P_d as a function of range for a fixed radial-velocity moving target

Clutter-limited Case. The foregoing discussion assumed that the target fell in the noise-limited (i.e., clutter-free) part of the doppler band. If the target falls in the sidelobe clutter region, the range performance will be degraded, since the total interference power (system noise plus clutter) against which the target must compete is increased. The foregoing discussion can be applied to the sidelobe clutter region, however, by interpreting R_o as the range where the signal is equal to sidelobe clutter plus system noise.[74–76] The CFAR loss may also be higher owing to the increased variability of the threshold when the clutter varies over the target detection region. More accurate calculations of detection performance in the sidelobe clutter limited case should include the proper clutter RCS fluctuation models and CFAR techniques.[77]

LIST OF ABBREVIATIONS

AESA	active electronically scanned array
A/D	analog-to-digital
AGC	automatic gain control
AM	amplitude modulation
CAGC	clutter automatic gain control
CFAR	constant false alarm rate
CNR	clutter-to-noise power ratio
CPI	coherent processing interval
CW	continuous wave
Δ_{AZ}	delta-azimuth antenna beam (used for monopulse angle estimation)
Δ_{EL}	delta-elevation antenna beam (used for monopulse angle estimation)
dBc	decibels with respect to the carrier
DC	direct current
DFT	discrete Fourier transform

DPD	digital product detector
ESA	electronically scanned array
FFT	fast Fourier transform
FM	frequency modulation
FM-CW	frequency-modulated continuous-wave
HRWS	high-PRF range-while-search
I	inphase
IF	intermediate frequency
INS	inertial navigation system
IPP	interpulse period
LNA	low-noise amplifier
LO	local oscillator
MF	matched filter
MRWS	medium-PRF range-while-search
MTI	moving target indication
MTT	multiple-target tracking
NAGC	noise automatic gain control
PAM	pulse-amplitude modulation
P_d	probability of detection
PC	pulse compression
PDI	postdetection integration (noncoherent integration)
P_{FA}	probability of false alarm
PM	phase modulation
PPM	pulse-position modulation
PRF	pulse repetition frequency
PWM	pulse-width modulation
Q	quadrature
RCS	radar cross section
RFI	radio frequency interference
rms	root-mean-square
RF	radio frequency
R/P	receiver protector
RWS	range-while-search
Σ	sum receive antenna beam (primary beam used for detection)
SLB	sidelobe blanker
SNR	signal-to-noise power ratio
STC	sensitivity time control
TWS	track-while-scan
T/R	transmit/receive
VS	velocity search

REFERENCES

1. *IEEE Standard Radar Definitions,* IEEE Std 686–1997, 1997, p. 20.
2. D. C. Schleher, *MTI and Pulsed Doppler Radar,* Norwood, MA: Artech House, Inc., 1991, pp. ix–x.
3. F. E. Nathanson, *Radar Design Principles,* 2nd Ed. New York: McGraw-Hill, 1991, pp. 471–472.
4. M. I. Skolnik, *Introduction to Radar Systems,* Chapter 3, 3rd Ed. New York: McGraw-Hill, 2000.
5. G. W. Stimson, *Introduction to Airborne Radar,* Chapter 3 & Part X, 2nd Ed., Raleigh, NC: SciTech Publishing, Inc., 1998.

6. P. Lacomme, J. Hardange, J. Marchais, and E. Normant, *Air and Spaceborne Radar Systems: An Introduction.,* Chapter 2, Norwich, NY: William Andrew Publishing, LLC, 2001.
7. S. A. Hovanessian, *Radar System Design and Analysis*, Chapter 12, Norwood, MA: Artech House, Inc., 1984.
8. M. I. Skolnik, *Radar Applications*, New York: IEEE Press, 1988.
9. R. J. Doviak, D. S. Zrnic, and D. S. Sirmans, "Doppler weather radar," in *Proceedings of the IEEE,* vol. 67, no. 11, 1979, pp. 1522–1553.
10. P. Mahapatra, *Aviation Weather Surveillance Systems: Advanced Radar and Surface Sensors for Flight Safety and Air Traffic Management,* London, UK: The Institution of Electrical Engineers, 1999.
11. K. C. Overman, K. A. Leahy, T. W. Lawrence, and R. J. Fritsch, "The future of surface surveillance-revolutionizing the view of the battlefield," in *Record of the IEEE 2000 International Radar Conference,* May 7–12, 2000, pp. 1–6.
12. Defense Science Board, Future DoD Airborne High-Frequency Radar Needs/Resources, Office of the Under Secretary of Defense for Acquisition and Technology, Washington, DC, April 2001.
13. M. I. Skolnik, *Introduction to Radar Systems,* 3rd Ed. New York: McGraw-Hill, 2002, pp. 171–172.
14. G. W. Stimson, *Introduction to Airborne Radar,* 2nd Ed. Raleigh, NC: SciTech Publishing, Inc., 1998, pp. 329–333.
15. F. C. Williams and M. E. Radant, "Airborne radar and the three PRFs," *Microwave Journal*, July 1983 and reprinted in M. I. Skolnik, *Radar Applications.* New York: IEEE Press, 1988, pp. 272–276.
16. D. C. Schleher, *MTI and Pulsed Doppler Radar*, Artech House, Inc., 1991, pp. 59–73.
17. G. Morris and L. Harkness, *Airborne Pulsed Doppler Radar,* 2nd Ed. Norwood, MA: Artech House, Inc., 1996, p. 4.
18. W. H. Long and K. A. Harriger, "Medium PRF for the AN/APG-66 radar," in *Proceedings of the IEEE,* vol. 73, issue 2, February 1985, pp. 301–311.
19. B. Cantrell, "ADC spurious signal mitigation in radar by modifying the LO," in *Proceedings of the 2004 IEEE Radar Conference,* April 26–29, 2004, pp. 151–156.
20. H. Hommel and H. Feldle, "Current status of airborne active phased array (AESA) radar systems and future trends," in *2005 IEEE MTT-S International Microwave Symposium Digest,* June 12–17, 2005, pp. 1449–1452.
21. S. M. Sherman, *Monopulse Principles and Techniques*, Norwood, MA: Artech House, Inc., 1984.
22. L. E. Pellon, "A double Nyquist digital product detector for quadrature sampling," *IEEE Transactions on Signal Processing*, vol. 40, issue 7, pp. 1670–1681, July 1992.
23. G. Minkler, *CFAR: The Principles of Automatic Radar Detection in Clutter*, Baltimore, MD: Magellan Book Company, 1990.
24. R. Nitzberg, *Radar Signal Processing and Adaptive Systems,* Chapter 7, Norwood, MA: Artech House, Inc., 1999.
25. M. Weiss, "Analysis of some modified cell-averaging CFAR processors in multiple-target situations," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-18, no.1, pp. 102–144, January 1982.
26. P. P. Gandhi and S. A. Kassam, "Analysis of CFAR processors in nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 24, no. 4, July 1988.
27. J. Farrell and R. Taylor, "Doppler radar clutter," *IEEE Transactions on Aeronautical & Navigational Electronics,* vol. ANE-11, pp. 162–172, September 1964 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section VI-2, Vol. 7. Norwood, MA: Artech House, Inc., 1978, pp. 351–361.
28. L. Helgostam and B. Ronnerstam, "Ground clutter calculation for airborne doppler radar," *IEEE Transactions on Military Electronics,* vol. MIL-9, pp. 294–297, July–October 1965.

29. A. L. Friedlander and L. J. Greenstein, "A generalized clutter computation procedure for airborne pulse doppler radars," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-6, pp. 51–61, January 1970 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section VI-3, Vol. 7, Norwood, MA: Artech House, Inc., 1978, pp. 363–374.
30. M. B. Ringel, "An advanced computer calculation of ground clutter in an airborne pulse doppler radar," in *NAECON '77 Record,* pp. 921–928 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section VI-4, Vol. 7. Norwood, MA: Artech House, Inc., 1978, pp. 375–382.
31. R. L. Mitchell, *Radar Signal Simulation,* Chapter 11, Norwood, MA: Artech House, Inc., 1976.
32. J. K. Jao and W. B. Goggins, "Efficient, closed-form computation of airborne pulse doppler clutter," in *Proceedings of the 1985 IEEE International Radar Conference,* Washington, DC, 1985, pp. 17–22.
33. W. A. Skillman, *SIGCLUT: Surface and Volumetric Clutter-to-Noise, Jammer and Target Signal-to-Noise Radar Calculation Software and User's Manual,* Norwood, MA: Artech House, Inc., 1987, pp. 1–4.
34. D. C. Schleher, *MTI and Doppler Radar,* Norwood, MA: Artech House, Inc., 1991, pp. 131–135.
35. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," in *Proceedings of the IEEE,* vol. 66, no. 1, January 1978, pp. 51–83.
36. W. A. Skillman, *Radar Calculations Using the TI-59 Programmable Calculator,* Norwood, MA: Artech House, Inc., 1983, p. 308.
37. R. E. Ziemer and J. A. Ziegler, "MTI improvement factors for weighted DFTs," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-16, pp. 393–397, May 1980.
38. H. R. Ward, "Doppler processor rejection of ambiguous clutter," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-11, July 1975 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section IV–11, Vol. 7. Norwood, MA: Artech House, Inc., 1978, pp. 299–301.
39. R. H. Fletcher and D. W. Burlage, "An initialization technique for improved MTI performance in phased array radar," in *Proceedings of the IEEE*, vol. 60, December 1972, pp. 1551–1552.
40. D. H. Harvey and T. L. Wood, "Design for sidelobe blanking systems," in *Proceedings of the 1980 IEEE International Radar Conference,* Washington, DC, 1980, pp. 410–416.
41. L. Maisel, "Performance of sidelobe blanking systems," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-4, pp. 174–180, March 1968.
42. H. M. Finn, R. S. Johnson, and P. Z. Peebles, "Fluctuating target detection in clutter using sidelobe blanking logic," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-7, pp. 147–159, May 1971.
43. A. Farina, *Antenna-based Signal Processing Techniques for Radar Systems*, Chapter 3, Norwood, MA: Artech House, Inc., 1992, pp. 59–93.
44. D. A. Shnidman and S. S. Toumodge, "Sidelobe blanking with integration and target fluctuation," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 38, no. 3, pp. 1023–1037, July 2002.
45. D. H. Mooney, "Post Detection STC in a Medium PRF Pulse Doppler Radar," U.S. Patent 4,095,222, June 13, 1978.
46. F. E. Nathanson, *Radar Design Principles,* 2nd Ed. New York: McGraw-Hill, Inc., 1991, pp. 281–282.
47. J. B. Tsui, *Digital Techniques for Wideband Receivers,* 2nd Ed., Raleigh, NC: SciTech Publishing Company, 2004, pp. 163–166.
48. L. P. Goetz and W. A. Skillman, "Master oscillator requirements for coherent radar sets," in *IEEE-NASA Symposium on Short Term Frequency Stability,* NASA-SP-80, November 1964.
49. R. S. Raven, "Requirements for master oscillators for coherent radar," in *Proceedings of the IEEE,* vol. 54, February 1966, pp. 237–243 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section V-I, Vol. 7, Artech House, Inc., Norwood, MA, 1978, pp. 317–323.
50. R. S. Raven, Correction to "Requirements for master oscillators for coherent radar," in *Proceedings of the IEEE,* vol. 55, issue 8, August 1967, p. 1425.

51. M. Gray, F. Hutchinson, D. Ridgely, F. Fruge, and D. Cooke, "Stability measurement problems and techniques for operational airborne pulse doppler radar," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-5, pp. 632–637, July 1969.
52. A. E. Acker, "Eliminating transmitted clutter in doppler radar systems," *Microwave Journal,* vol. 18, pp. 47–50, November 1975 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section V-3, Vol. 7. Norwood, MA: Artech House, Inc., 1978, pp. 331–336.
53. J. A. Scheer and J. L. Kurtz, *Coherent Radar Performance Estimation,* Norwood, MA: Artech House, Inc., 1993, pp. 158–159.
54. S. J. Goldman, *Phase Noise Analysis in Radar Systems Using Personal Computers,* Chapter 2, New York: John Wiley & Sons, Inc, 1989.
55. G. V. Trunk and M. W. Kim, "Ambiguity resolution of multiple targets using pulse-doppler waveforms," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 30, no. 4, pp. 1130–1137, October 1994.
56. F. E. Nathanson, *Radar Design Principles,* 2nd Ed. New York: McGraw-Hill, Inc., 1991, pp. 449–452.
57. M. B. Ringel, "The effect of linear FM on the ground clutter in an airborne pulse doppler radar," in *NAECON '79 Record,* vol. 2, Dayton, OH, May 15–17, 1979, pp. 790–795.
58. F. E. Nathanson, *Radar Design Principles,* 2nd Ed. New York: McGraw-Hill, Inc., 1991, pp. 120–123.
59. G. W. Stimson, *Introduction to Airborne Radar,* 2nd Ed. Mendham, NJ: SciTech Publishing, Inc., 1998, pp. 506–507.
60. P. L. Bogler, *Radar Principles with Applications to Tracking Systems,* New York: John Wiley & Sons, Inc., 1990, pp. 262–266.
61. R. A. Dana and D. Moraitis, "Probability of detecting a Swerling I target on two correlated observations," *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-17, no. 5, pp. 727–730, September 1981.
62. R. E. Ziemer, T. Lewis, and L. Guthrie, "Degradation analysis of pulse doppler radars due to signal processing," in *NAECON 1977 Record,* pp. 938–945 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section IV-12, Vol. 7, Norwood, MA: Artech House, Inc., 1978, pp. 303–312.
63. P. Lacomme, J. Hardange, J. Marchais, and E. Normant, *Air and Spaceborne Radar Systems: An Introduction,* Norwich, NY: William Andrew Publishing, LLC, 2001, pp. 150–151.
64. J. I. Marcum, "A statistical theory of target detection by pulsed radar," *IEEE Transactions on Information Theory,* vol. IT-6, pp. 59–267, April 1960.
65. P. Swerling, "Probability of detection for fluctuating targets," *IEEE Transactions on Information Theory,* vol. IT-6, pp. 269–308, April 1960.
66. L. F. Fehlner, "Target detection by a pulsed radar," Report TG 451, Johns Hopkins University, Applied Physics Laboratory, Laurel, MD, 2 July 1962.
67. D. P. Meyer and H. A. Mayer, *Radar Target Detection: Handbook of Theory and Practice.* New York: Academic Press, 1973.
68. J. V. DiFranco and W. L. Rubin, *Radar Detection,* Norwood, MA: Artech House, Inc., 1980.
69. J. V. DiFranco and W. L. Rubin, *Radar Detection,* Norwood, MA: Artech House, Inc., 1980, pp. 287–445.
70. D. A. Shnidman, "Determination of required SNR values," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 38, no. 3, pp. 1059–1064, July 2002.
71. D. K. Barton, "Universal equations for radar target detection," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 41, no. 3, pp. 1049–1052, July 2005.
72. M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions,* 3rd Ed. New York, John Wiley & Sons, Inc., 2000, p. 13.
73. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* 2nd Ed. Cambridge, UK: Cambridge University Press, 1992, pp. 213–222.

74. D. Mooney and G. Ralston, "Performance in clutter of airborne pulse MTI, CW doppler and pulse doppler radar," in *1961 IRE Convention Record*, vol. 9, part 5, 1961, pp. 55–62 and reprinted in D. K. Barton, *CW and Doppler Radars,* Section VI-1, Vol. 7. Norwood, MA: Artech House, Inc., 1978, pp. 343–350.

75. M. B. Ringel, "Detection range analysis of an airborne medium PRF radar," in *IEEE 1981 NAECON Record,* Dayton, OH, 1981, pp. 358–362.

76. P. E. Holbourn and A. M. Kinghorn, "Performance analysis of airborne pulse doppler radar," in *Proceedings of the 1985 IEEE International Radar Conference,* Washington, DC, 1985, pp. 12–16.

77. D. A. Shnidman, "Radar detection probabilities and their calculation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 3, pp. 928–950, July 1995.

Chapter 5
Multifunctional Radar Systems for Fighter Aircraft

David Lynch, Jr.

DL Sciences, Inc.

Carlo Kopp

Monash University

5.1 INTRODUCTION

In spite of more than a half century of improvements in radar performance and reliability, the effort required for deployment, operation, and maintenance of most radars is substantial. Furthermore, the power-aperture product is never as large as desired. The forward projected area as well as avionics weight is very costly in most fighter aircraft parameters. These parameters have motivated users, buyers, and designers to want more functions in a single radar and its complementary processing suite. As a result, most modern fighter radars are multifunctional—providing radar, navigation, landing aids, data link, and Electronic Counter Measures (ECM) functions.[1,2] The primary enabler for multifunctional radar is software-defined signal and data processing, first introduced in the mid 1970s.[3–6] Software programmability allows many radar system modes to be performed using the same RF hardware. In addition, modern navigation aids work so well that each radar mode is defined by its earth situation-geometry with almost all waveform parameters set by local earth conditions.[7,9] The modern radar often is net-centric, using and providing data to a communications network and where suitably equipped, has its own Internet protocol (IP) address.

Multifunctionality is not dependent on antenna type. In fact, the mechanically scanned AN/APG-65, 70, and 73 radars have demonstrated multifunctionality in combat.[7] However, multifunctionality is facilitated by Active Electronically Scanned Antenna (AESA) arrays. The multifunctional AESA radar in the F/A-18E/F fighter is shown with a protective cover over the array in Figure 5.1. The AESA is shaped and canted upward to aid in some modes and to minimize reflections to enemy radars.[8]

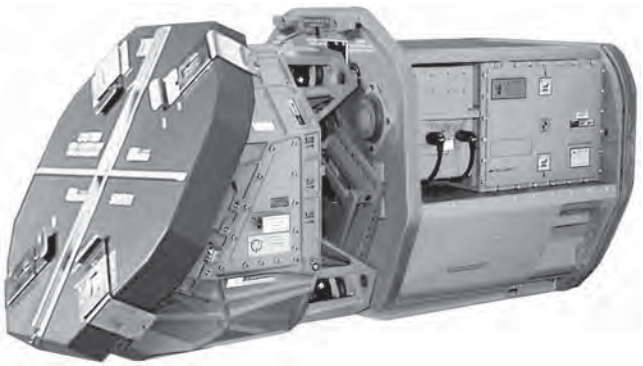


FIGURE 5.1 AN/APG-79 Multifunctional AESA Radar (*Courtesy Raytheon Company*[12])

This chapter addresses *what* signals are emitted and *why* they are needed in a Multifunctional Fighter Aircraft Radar (MFAR). The *why* begins with typical missions, which shows the geometry that gives rise to each radar mode and waveform, lists representative radar modes, and shows typical modern airborne radar mode interleaving and timing. The answer to *what* is provided by typical waveform variations and a few examples. The examples are not from any single radar but are a composite of modern radars. The general MFAR idea is illustrated in Figure 5.2. It shows time multiplexed operations for air-to-air (A-A), air-to-surface (A-S), electronic warfare (EW), and communication from the same radio frequency (RF) hardware and processing complex often over most of the microwave band.[9,11] Sometimes, multiple functions can be performed simultaneously if a common waveform is used.

The antenna aperture usually has multiple phase centers enabling measurement for Space-Time Adaptive Processing (STAP),[13] Displaced Phase Center Antenna (DPCA)

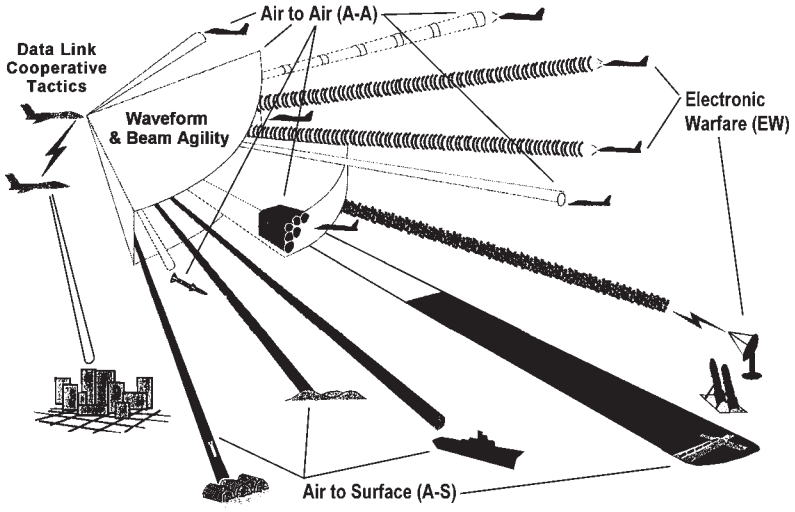


FIGURE 5.2 MFAR interleaves A-S, A-A, and EW functions (*adapted*[9])

processing, conventional monopulse angle tracking, jammer nulling, and out-of-band angle-of-arrival (AOA) estimation. The optimum placement of phase centers is an important design tradeoff. A phase center is an antenna aperture channel, which is offset in space and provides a partially or fully independent measurement of an incoming electromagnetic wavefront. For example, a one-dimensional phase monopulse has two phase centers, a two-dimensional phase monopulse has four phase centers, DPCA has two or more phase centers, a radar with a guard horn for sidelobe suppression has two phase centers, and an adaptive array may have many phase centers.[13–16] STAP is an extension of the classic theory for a matched filter in the presence of non–white noise, which includes both time and space.

Overall weapon system requirements usually favor *X* or K_u band for the operating frequency of a MFAR. In addition, the MFAR apertures and associated transmitter are usually the largest on an aircraft and hence, can create the highest Effective Radiated Power (ERP) for jamming adversary radars and data links, where these are in-band.

Multifunctional Radar Architecture. An example MFAR block diagram is shown in Figure 5.3. The modern integrated avionic suite concept blurs the boundaries between traditional radar functions and other sensors, countermeasures, weapons, and communications (see Figures 5.12 and 5.14 later in the chapter). There is a microwave and RF suite; an electro-optical, infrared, ultraviolet (EO) suite; a stores management suite; a controls and displays suite; a multiply-redundant vehicle management suite, and a multiply-redundant processor complex.

Each microwave and/or RF aperture may have some embedded signal conditioning but then may be multiplexed to standardized common design RF, filter, frequency reference, analog to digital conversion (A/D), input-output (I/O), and control modules. A similar design concept is used for the electro-optical (EO) sensors, stores management,

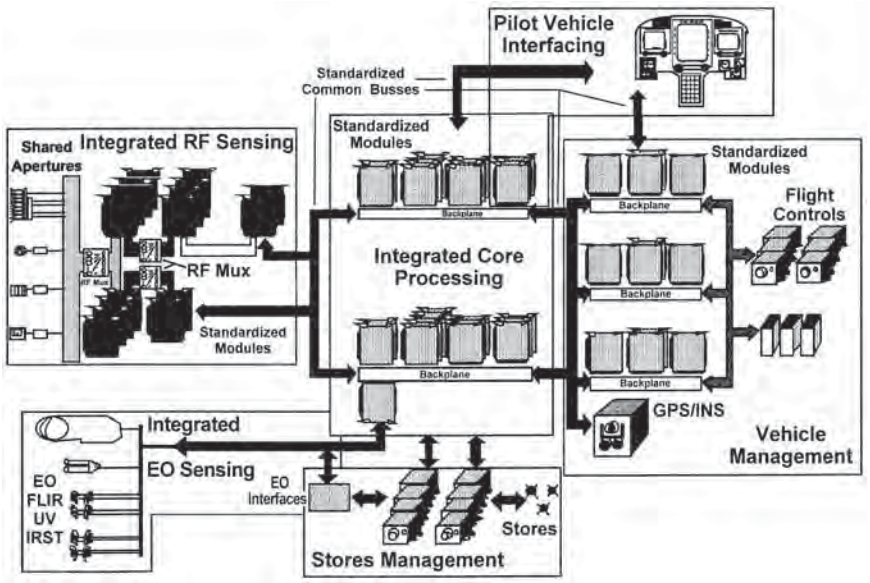


FIGURE 5.3 MFAR merged with other sensors (*adapted*[2])

vehicle management, pilot-vehicle interface, and integrated core-processing suite. There is substantial data traffic between the core processing and the sensors to provide pointing, cueing, tracking, and multisensor fusion of detections. The aim of this approach is to provide a shared pool of computational resources, which may be flexibly allocated between sensors and functions.

The sensors may contain dedicated motion sensing, but long-term navigation is provided by the vehicle management global positioning system and inertial navigation system (GPS/INS). The on-radar motion sensing must sense position to a fraction of the transmitted wavelength over the coherent processing interval. This is usually done with inertial sensors such as accelerometers and gyros with very high sampling rates. An inertial navigation system estimates the position of the aircraft in a worldwide coordinate space by integrating the outputs of the gyros and accelerometers, typically using Kalman filtering techniques. Accumulated errors in such a system can be corrected by using GPS updates as well as known reference points measured with the radar, or EO sensors.

There may be dozens or hundreds of stored program devices distributed throughout the avionics. These lower level functional suites are connected by standardized busses, which may be fiber optic or wired. The programmable devices are controlled by software operating environments invoking programs. The architecture objective is to have standard interfaces, few unique assemblies, and single-level maintenance.

The suite of microwave and RF apertures in a fighter aircraft might appear as shown in Figure 5.4. As many as 20 apertures may be distributed throughout the vehicle, performing radar, data link, navigation, missile warning, direction finding, jamming, or other functions over a frequency range covering several decades.[2] There are apertures distributed over the aircraft that point forward and aft, right and left, as well as up and down. Some apertures will be shared for communications, radio navigation, and identification (CNI) as well as identification, friend or foe (IFF) due to compatible frequencies and geometries. Data links such as JTIDS/Link 16 and Link 22 can share apertures with GPS and L band satellite communications (L SATCOM). EW apertures must be broadband by nature and can be shared with radar warning receivers (RWR), radar auxiliaries, and some types of CNIs.

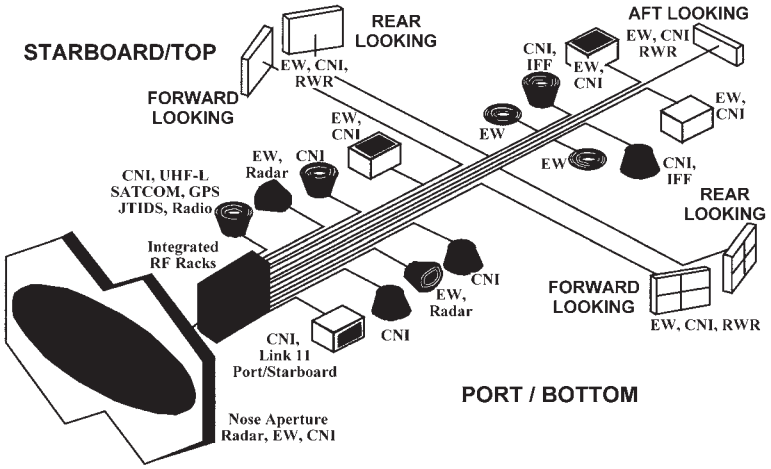


FIGURE 5.4 MFAR RF apertures share low-level RF (*adapted*[2]).

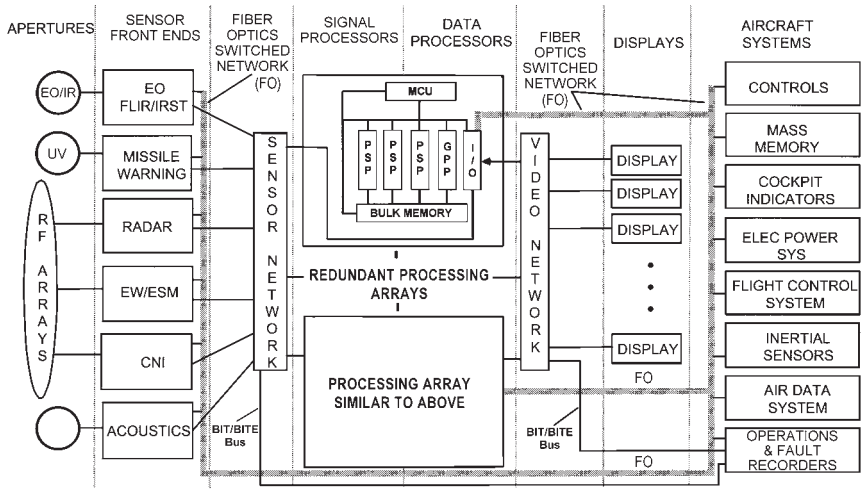


FIGURE 5.5 MFAR processing (*adapted*[2])

The apertures are signal conditioned, controlled, and interfaced through busses in the aircraft with remaining processing performed either in a common processor complex, as shown in Figure 5.5, or in federated processors distributed throughout the aircraft. One important class of standardized modules contains basic timing and programmable event generators (PEG) that create accurate timing for Pulse Repetition Frequencies (PRFs), analog to digital conversion (A/D) sampling, pulse and chip widths, blanking gates, beam repointing commands, and other synchronized real-time interrupts. A second class contains RF and intermediate frequency (IF) amplification and mixing. A third class contains low noise frequency synthesizers, which may include Direct Digital-frequency Synthesis (DDS). A/D converters and control interface modules are the final class. Bussing protocols and speeds must have adequate reserves to insure fail-safe real-time operation.

The functional block diagram and operation of a specific sensor mode is then overlaid on this hardware and software infrastructure. A specific mode is implemented in an applications program in the same sense that word-processing is on a personal computer (PC). Carrying the analogy further, common experience with the unreliability of PC hardware and software requires that a system of the type depicted in Figure 5.3 must be redundant, error checking, trusted, fail safe in the presence of faults, and embody strict program execution security. This is a very challenging system engineering task. Exhaustive mathematical assurance and system testing is required, which is completely different from current commercial personal computer practice.

A notional MFAR integrated core processing complex with its corresponding interfaces similar to that shown in Figure 5.3 is shown in Figure 5.5, where there are multiple redundant processing arrays that contain standardized modules connected in a non-blocking switched network. Internal and external busses connect the individual processing arrays to each other as well as to the other suites, sensors, controls, and displays.

Usually, there are both parallel electrical signal busses as well as serial fiber optic busses depending on speed and total length in the aircraft.[2] The signal and data processor complex contains multiple processor and memory entities, which might be

on a single chip or on separate chips depending on yield, complexity, speed, cache size, and so on. Each processor array may consist of programmable signal processors (PSP), general purpose processors (GPP), bulk memory (BM), input-output (I/O), and a master control unit (MCU). The PSPs perform signal processing on arrays of sensor data. The GPPs perform processing in which there are large numbers of conditional branches. The MCU issues programs to PSPs, GPPs, and BM, as well as manages overall execution and control. Typical processing speed is 6000 MIPS (millions of instructions per second) per chip but might be 32 GIPS (billions of instructions per second) in the near future.[17] Clock frequencies are limited by on-chip signal propagation but are up to 4 GHz (gigahertz) and could be 10 GHz in the near future.[18] Sensor processing has arrived at the point where the conception of successful algorithms is more important than the computational horsepower necessary to carry them out.

MFAR Software Structure. Improper operation of many fighter systems can be hazardous. As previously mentioned, the software must be exhaustively tested, error checked, mathematically trusted, failsafe in the presence of faults, and embody strict program execution security. One of the most important aspects is rigid adherence to a structured program architecture. An object-based hierarchical structure, where each level is subordinate to the level above and subprograms are called in strict sequence, is necessary. It also requires, among other things, that subprograms never call themselves (recursive code) or any others at their execution level. Subprograms (objects) are called, receive execution parameters from the level above (the parent), and return results back to the calling level.[94] An example of such a software structure is shown in Figures 5.6 and 5.7. The software would be executed in the hardware shown in Figure 5.5.

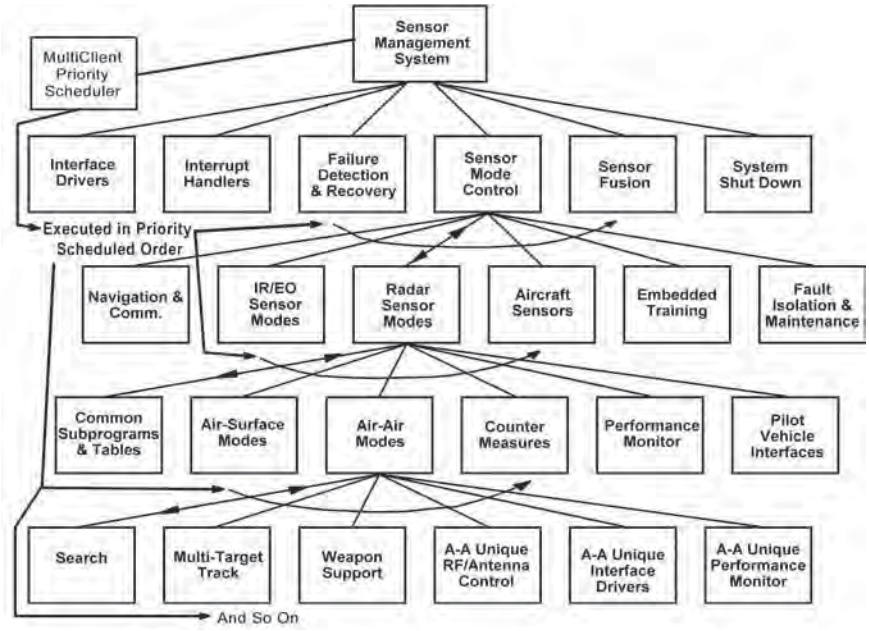


FIGURE 5.6 MFAR structured software

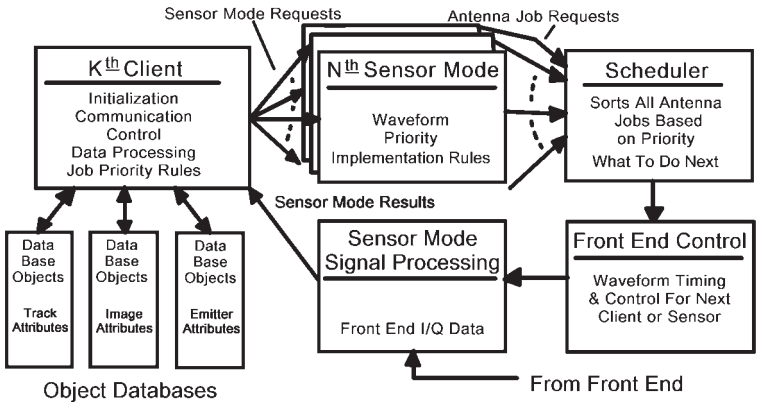


FIGURE 5.7 MFAR priority scheduling

An MFAR can support many activities (or modes) concurrently by interleaving their respective data collections. Surveillance, track updates, and ground maps are examples of such activities. The software needed to support each activity is mapped to a specific client module, as shown in Figure 5.7. Each client module is responsible for maintaining its own object database and for requesting use of the aperture. Requests are made by submitting antenna job requests that specify both the waveform to be used (how to do it) and the priority and urgency of the request.

A scheduler executes during each data collection interval and decides what to do next, based on the priorities and urgencies of the antenna job requests that have been received. This keeps the aperture busy and responsive to the latest activity requests. Following the selection of the antenna job by the scheduler, the front-end (transmit and receive) hardware is configured, and in-phase and quadriture (I/Q) data is collected and sent to the signal processors. There, the data is processed in a manner defined by the sensor mode, and the signal processing results are returned to the client that requested them. This typically results in database updates and/or new antenna job requests from the client. New activities can be added at any time using this modular approach.

Although this structure is complex and the software encompasses millions of lines of code, modern MFAR software integrity can be maintained with strict control of interfaces, formal configuration management processes, and formal verification and validation software tools. In addition, most subprograms are driven by read-only tables, as shown in Figure 5.7, so that the evolution of aircraft tactics, capabilities, and hardware do not require rewrites of validated subprograms. Software versions (builds) are updated every year throughout the lifetime of the system, which may be decades. Each subprogram must have table driven error checking as well. Many lower levels are not shown in Figures 5.6 and 5.7; there may be several thousand subprograms in all.

Range Doppler Situation. Modern radars have the luxury of interleaving most of the modes suggested in Figure 5.2 in real time and selecting the best available time or aircraft position to invoke each mode as the mission requires.[7,9]

The geometry that must be solved each time is shown in Figure 5.8. The fighter aircraft pulse doppler geometry is centered around the aircraft traveling at a velocity, V_a, and at an altitude, h, above the Earth's surface. The radar pulse repetition frequency

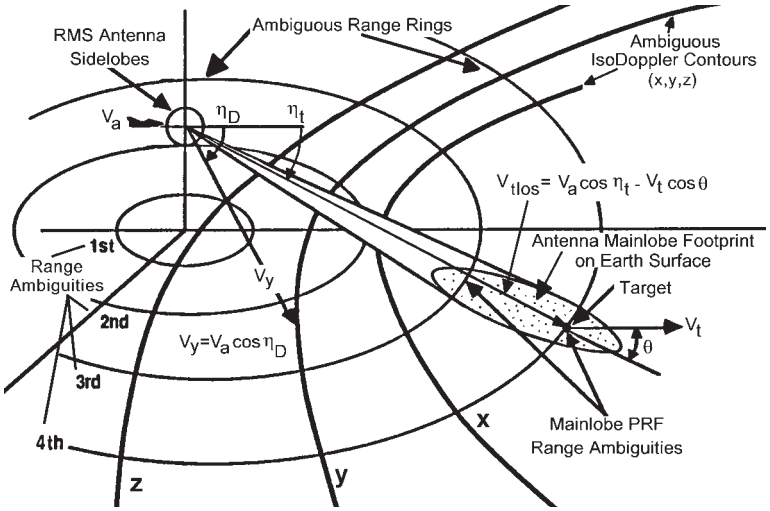


FIGURE 5.8 Strike fighter pulse doppler geometry[7,9]

(PRF) gives rise to a series of range (1, 2, 3, 4) and doppler (x, y, z) ambiguities, as shown in Figure 5.8, which intercept the Earth's surface as range "rings" and isodoppler "hyperbolas" (because the Earth is a rough geoid, constant range and doppler contours are not actually rings or hyperbolas). The radar antenna pattern intercepts the limb of the Earth usually in both the main beam and sidelobes. A target in the main beam at range, R_t, and velocity, V_t, may have to be observed in the presence of both range and doppler ambiguities. Only the target's line-of-sight velocity, V_{tlos}, is observable on a short term basis. The radar designers' problem is to select the best waveform in this target-clutter geometry. Historically, these waveforms were selected ahead of time and built into the radar hardware and software. Most modern airborne radars solve this geometry in real time and continuously select the best available frequency, PRF, pulsewidth, transmit power, scan pattern, etc.

Unfortunately, the specifics of the waveform are unpredictable even to the radar without exact knowledge of the aircraft-target-earth velocity-geometry set and mode of operation requested by the operator or mission software. This makes testing quite difficult; fortunately, test equipment has come a long way. Hardware-in-the-loop testing using real-time simulation of the entire geometry and external world in the radar integration laboratory is commonly employed.

Active Electronically Scanned Array (AESA). Although multifunctional radars have been deployed with mechanically scanned and electronically scanned antennas, fully multifunctional radars use Active Electronically Scanned Arrays (AESA), which contain a transmit-receive channel (T/R) for each radiator.[1] The advantages of AESA are fast adaptive beam shaping and agility, improved power efficiency, improved mode interleaving, simultaneous multiple weapon support, and reduced observability.[19–23] Perhaps half the cost and complexity of an AESA is in the T/R channels. That said, however, the feed network, beam steering controller (BSC), AESA power supply, and cooling subsystem (air or liquid) are equally important.[9,11]

A major enabler for AESAs is the state of the art in microwave integrated circuits.[1] This has followed the dramatic cost and performance gains available in most semiconductor technologies. Each T/R channel has self-diagnosis features, which can detect failure and communicate that to the beam steering controller for failure compensation. AESAs can accommodate up to 10% failures with very little degradation if properly compensated in the BSC.[24]

From an MFAR point of view, the important parameters are volumetric densities high enough to support less than 1/2 wavelength spacing; radiated power densities high enough to support 4 watts per sq. cm.; radiated-to-prime-power efficiencies greater than 25%; bandwidth of several GHz on transmit and almost twice that bandwidth on receive; phase and amplitude calibration and control adequate to provide at least –50 dB rms sidelobes; amplitude control adequate to provide 50 dB power management; noise performance adequate to support the subclutter visibility requirements; and finally, sufficient storage and computing to allow beam repointing/adjustment in a fraction of 1 msec. Fast beam adjustment requires high-speed busses to each T/R channel.

One of the principal advantages of an AESA is the ability to manage both power and spatial coverage on a short-term basis (10s of msec.). Often another advantage is that both the noise figure is lower and radiated power is higher for a given amount of prime power. This is because the RF path lengths can be much shorter, which usually leads to lower front-end losses. Each radiating element is usually designed to be very broadband and is driven by a T/R channel in a typical AESA array. There are typically a few thousand channels in an MFAR AESA. Each channel contains first-level power regulation, filtering, logic, calibration tables as well as the obvious RF functions. Some channels in the array are dedicated to other functions such as calibration, jammer nulling, sidelobe blanking, close in missile datalink, out-of-band direction finding, etc.[19,25,26,27] Also, there are usually some channels at the edge of the array that are passive and improve the sidelobes and RCS pattern.[8]

Figure 5.9 shows the comparison between a conventional mechanically scanned radar with the low-noise amplifier and a high-power traveling wave tube transmitter mounted off the gimbal versus a real-time adapted AESA with two different scan regimes for the same amount of input prime power. AESA performance falls off for large scan coverage because of the lower projected aperture area for a fixed mounting as shown in Figure 5.1. A mechanical scan has the same projected area in all directions and large scan angles marginally reduce radome losses, which results in slightly improved large angle performance. Nonetheless, AESA performance is usually superior inside

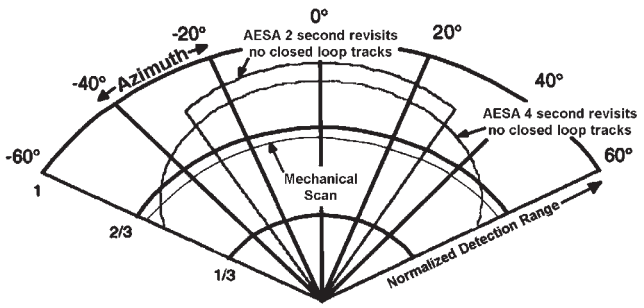


FIGURE 5.9 Example AESA management comparison (*adapted*[9])

a ±50° azimuth scan.[9,11,28] Usually, a fighter can't engage at long range outside this azimuth for kinematic reasons.

The performance differences depicted in Figure 5.9 are the result of three factors: the installed aperture can be 20% larger in net projected area at the aircraft in-flight horizontal due to elimination of gimbal swing space, 2:1 higher radiated power due to lower losses and better efficiency, and 60% lower losses before the low-noise amplifier. The other major advantage is that search volume can be changed dynamically to fit the instant tactical situation, as suggested in Figure 5.9.[28]

The feed network is mundane but critically important. In single tube transmitters, the feed is heavy because it must carry high power at low loss. AESA feeds use smaller coax, stripline, microstrip, or RF modulated light in fiber optics for transmit and receive RF, since less than 10 watts RF or optical is usually required. However, significant DC power is still required for RF feed distribution amplifiers because thousands must be driven. Cost, weight, and complexity is still an issue because multiple phase centers necessary for adaptive array performance require multiple manifolds. Usually, once a subarray is formed in the manifolds, it is digitized and multiplexed for adaptive signal processing.

Another important function is beam steering control (BSC). The BSC does array calibration, failed element compensation,[8,24] phase and amplitude setting for beam steering as well as space-time adaptive operation.[29–33] The BSC is usually realized with a combination of general purpose processing of the type found in a personal computer with very high speed incremental phase and amplitude calculation and T/R module interface hardware. Both scanning and adaptive operation require very low latency (i.e., the time between the sensed need and the first pulse at the target is usually 1 msec) beam control in a high-speed aircraft platform.

Lastly, the AESA requires a very significant power supply.[1] Power supplies have a history of being heavy, hot, and unreliable. Even the best systems still have overall power efficiencies (prime power in to RF out in space) in the 10–25% region in spite of years of development. The typical AESA requires low voltage and high current at the T/R channel. This forces large conductors in the absence of high power lightweight superconductors (not available at this writing). It also requires very low voltage drop rectifiers and regulators. Cooling is generally a significant performance burden. Usually, the power supplies are distributed to improve reliability and fault tolerance. Often, power converters are operated at switching frequencies up to several hundred megahertz to reduce the size of magnetics and filter components, and sometimes, the switching frequencies are synchronized to the radar master clock.

5.2 TYPICAL MISSIONS AND MODES

Air-to-Surface Mission Profile. The mode structure of any modern fighter aircraft arises from mission profiles.[7,9] One typical mission profile for an air-to-surface (A-S) strike is shown in Figure 5.10. The mission profile begins with a takeoff, continues through flight to a target, and ultimately returns to the starting point. Along the way, the aircraft uses a variety of modes to navigate, search and acquire targets, track targets, deliver weapons, assess battle damage, engage in countermeasures, and monitor and calibrate its performance. AESAs have demonstrated simultaneous multiple weapon deliveries.[22]

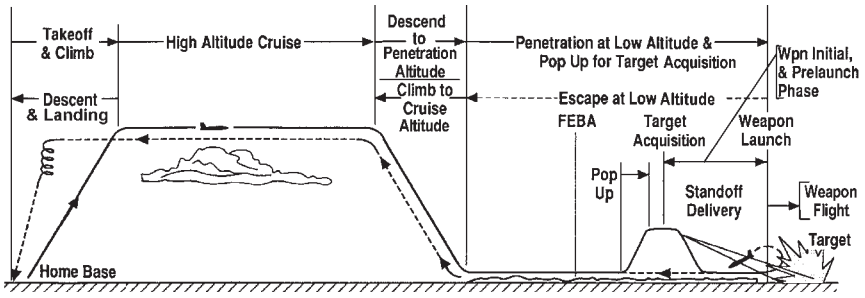


FIGURE 5.10 Typical air-to-surface mission profile[7,9]

Air-to-Surface Mode Suite. The mission naturally creates the need for an air-to-surface mode suite[7,9] for fighter radar, as shown in Figure 5.11. Each general category of operation contains modes primarily for that function, but modes will often be invoked during other parts of the mission. Within each mode shown in Figure 5.11, there is optimization for the particular combination of altitude, range to the target, antenna footprint on the Earth's surface, relative target and clutter doppler, dwell time available, predicted target statistical behavior, transmitted frequency, and desired resolution. Obviously, each mode must not compromise some required level of mission stealth.[34–37] A modern fighter is net-centric and exchanges substantial information with other systems. Both the fighter's wingman, support aircraft, and surface nodes may exchange complete data and tasking in real time to facilitate a mission. The fighter and its wingman will coordinate mode tasking so that during a high resolution ground map, which could take a minute to form, the wingman might be performing an air-to-air search and track to protect both of them.

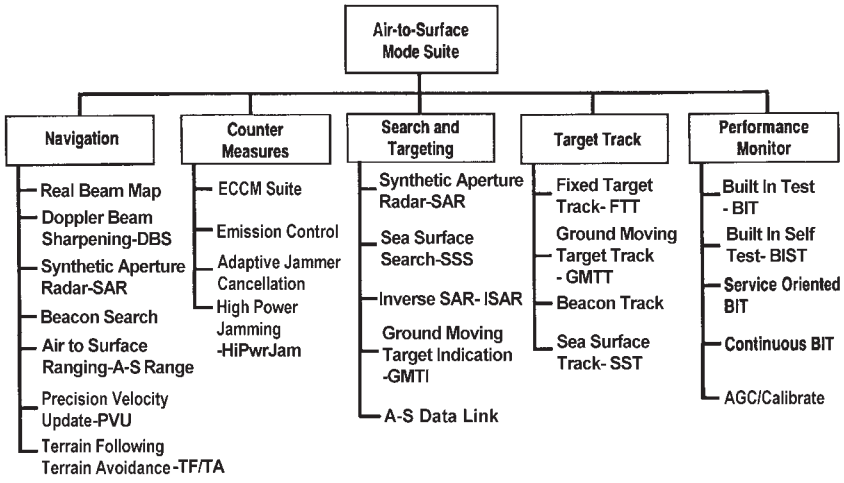


FIGURE 5.11 Fighter aircraft air-to-surface radar mode suite[7,9]

Some modes are used for several operational categories, such as real beam map (RBM), fixed target track (FTT), doppler beam sharpening (DBS), and synthetic aperture radar (SAR), used not only for navigation but also for acquisition and weapon delivery to fixed targets.[38–43] SAR may also be used to detect targets in earthworks or trenches covered with canvas and a small amount of dirt, which are invisible to EO or IR sensors. Similarly, air-to-surface ranging (A-S Range) and precision velocity update (PVU) may be used for weapon support to improve delivery accuracy as well as navigation.[7,9]

Terrain following and terrain avoidance (TF/TA) is used for navigation at very low altitudes or in mountainous terrain. Sea surface search (SSS), sea surface track (SST), and inverse synthetic aperture radar (ISAR), which will be described later in the chapter, are used primarily for the acquisition and recognition of ship targets. Ground moving target indication (GMTI) and ground moving target tracking (GMTT) are used primarily for the acquisition and recognition of surface vehicle targets but also for recognizing large movements of soldiers and materials in a battle-space. High power jamming (HiPwrJam) is a countermeasure available from AESAs due to their natural broadband, beam agile, high gain, and high power attributes. AESAs also allow long range air-to-surface data links (A-S Data Link) through the radar primarily for map imagery. Because there may be thousands of wavelengths and a gain of millions through a radar, automatic gain control and calibration (AGC/CAL) is usually required fairly often. Modes optimized for this function are invoked throughout a mission.

Waveform Variations by Mode. Although the specific waveform is hard to predict, typical waveform variations can be tabulated based on observed behavior of a number of existing A-S radar systems. Table 5.1 shows the range of parameters that can be observed as a function of radar mode. The parameter ranges listed are PRF, pulse width, duty cycle, pulse compression ratio, independent frequency looks, pulses per coherent processing interval (CPI), transmitted bandwidth, and total pulses in a Time-On-Target (T_{OT}).

Obviously, most radars do not contain all of this variation, but modes exist in many fighter aircraft, which represent a good fraction of the parameter range. Most fighter radars are frequency agile since they will be operated in close proximity to similar or identical systems. The frequency usually changes in a carefully controlled, completely coherent manner during a CPI.[8] This can be a weakness for certain kinds of jamming since the phase and frequency of the next pulse is predictable. Sometimes to counteract this weakness, the frequency sequence is pseudorandom from a predetermined set with known autocorrelation properties, for example, Frank, Costas, Viterbi, P codes.[16] A major difficulty with complex wideband frequency coding is that the phase shifters in a phase scanned array must be changed on an intra- or inter-pulse basis greatly complicating beam steering control and absolute T/R channel phase delay. Another challenge is minimizing power supply phase pulling when PRFs and pulsewidths vary over more than 100:1 range. MFAR systems not only have a wide variation in PRF and pulsewidth but also usually exhibit large instant and total bandwidth. Coupled with the large bandwidth is the requirement for long coherent integration times. This requirement naturally leads to extreme stability master oscillators and ultra low-noise synthesizers.[44]

Air-to-Air Mission Profile. Just as with an air-to-surface mission, the mode structure of a modern fighter aircraft air-to-air mission arises from its profile.[45] A typical mission profile for air-to-air (A-A) is shown in Figure 5.12. The mission profile

TABLE 5.1 Typical Waveform Parameters A-S Modes[7,9]

Radar Modes	PRF (kHz)	Pulse Width (μsec)	Duty Cycle (%)	Pulse Comp. Ratio	Freq. Looks	Pulses Per CPI	Transmitted Bandwidth (MHz)	Total Pulses in T_{OT}
Real-Beam Map	0.5–2	1–200	0.1–10	1–200	1–4	1–8	0.2–10	8–100
Doppler Beam Sharp.	1–4	1–60	0.3–25	13–256	1–4	20–800	5–25	20–1.6k
SAR	1–10	3–60	1–25	32–16384	1–4	70–20k	10–500	150–100k
A-S Range	1–8	0.1–10	0.1–10	1–256	4–8	1–8	1–50	5–100
PVU	2–100	1–25	0.01–25	1–16	4–32	20–1024	1–10	5–1000
TF/TA	2–20	0.1–10	0.05–5	1–32	16–64	1–8	3–15	20–60
Sea-Surface Search	0.5–2	1–200	0.1–10	1–20000	4–32	1–8	0.2–500	8–100
Inverse SAR	1–25	1–60	0.1–10	13–256	1–4	20–256	5–100	20–1000
GMTI	3–10	2–60	0.1–25	1–256	1–4	20–256	0.5–15	20–550
Fixed Target Track	2–20	0.1–10	0.1–10	1–256	4–8	1–8	1–50	20–1000
GMTT	2–16	2–60	0.1–25	1–256	1–4	20–256	0.5–15	20–1000
Sea-Surface Track	2–20	1–200	0.1–10	1–200	1–4	20–256	0.2–10	20–1000
HiPwr Jam	50–300	3–10	10–50	13–512	1–8	1–8	1–100	200–2k
Cal/A.G.C. A-S Data Link	0.5–20 8–300	0.1–200 0.8–20	0.01–50 1–100	1–16384 13–32768	1–8 1–75	8–64 100–500	0.2–500 0.5–250	8–64 1.3k–80k

begins with an airfield or carrier takeoff, continues through flight penetrating into an enemy battle-space, searches for air targets to attack, and ultimately returns to the starting point. Along the way, the aircraft uses a variety of modes to navigate; exchange data with command, control, communications, intelligence, surveillance,

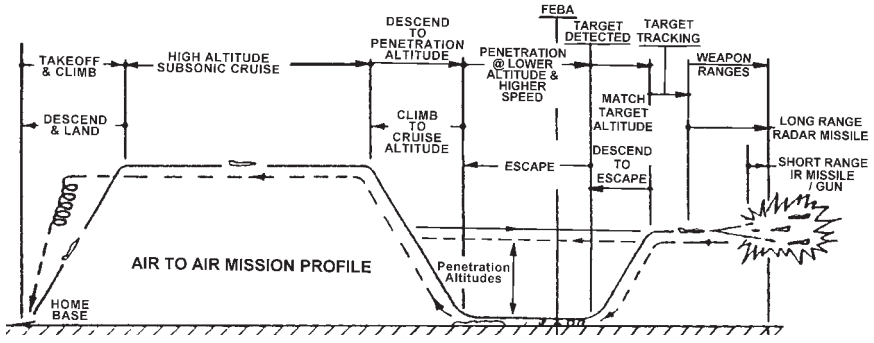


FIGURE 5.12 Typical A-A mission profile[45]

reconnaissance (C3ISR) assets; search and acquire airborne targets; track and separate benign targets from threats; deliver weapons; escape and engage in countermeasures; monitor and calibrate its performance, and return to base.

Air-to-Air Mode Suite. Similarly, the A-A mission naturally creates the need for a corresponding mode suite for the radar, as shown in Figure 5.13.[46,47] At the radar sensor and air-to-air mode software level, there is adaptive task prioritization to insure that the highest processor prioritized, pilot-selected threat is serviced first. Passive modes are interleaved with active operation to improve survivability and passive tracking and ID. Each mode shown in Figure 5.13 is optimized in real time for the particular combination of altitude, range to the target, density of target threats, antenna footprint on the Earth's surface, relative target and clutter doppler, dwell time available, predicted target statistical behavior, transmitted frequency, and desired resolution.[9,11]

The mode category "autonomous and cued search" contains the modes most commonly associated with fighter radars. There are usually two range-gated high pulse repetition frequency (HPRF) modes: velocity search (VS) primarily dedicated to longest range detection and range while search (RWS), which uses some form of FM ranging to estimate target range. There is a medium PRF (MPRF) mode, which provides all aspect velocity-range search (VRS) at the expense of poorer long-range performance. In addition, there are two passive modes: passive search and ranging, in which the radar detects and estimates range and angle to an emitter or bistatically (wingman or support aircraft) illuminated target and ESM shared aperture in which the RF and processor complex detects, estimates waveform parameters, and records them for future use. Passive search may be combined with cued burst ranging to better estimate emitter location. Extended volume search is a mode used with cueing from another on- or off-board sensor in what normally would be an unfavorable geometry.[85]

Many modes and functions are shared in common with A-S, especially countermeasures and performance monitoring. Extremely important in both modes is implementation of emissions control to minimize the ability of the adversary to detect, track, and attack using the radar emission.[16] Without care, these emissions can easily serve as a strong guidance signal for a hostile antiradiation missile (ARM).[50,51] Antenna apertures that have multiple independent phase centers can perform both adaptive clutter cancellation as well as jammer cancellation with suitable hardware and software.[14,27,29–33]

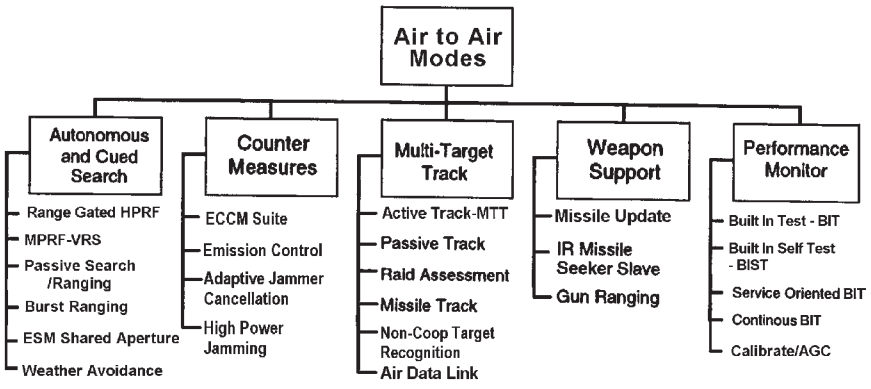


FIGURE 5.13 A-A mode suite[9]

The subsuite of multi-target track (MTT) contains conventional track while scan (TWS), passive tracking of emitters or echoes from bistatic illumination, missile tracking with or without a missile datalink or beacon, and several modes to recognize target number and type: raid assessment and noncooperative target recognition (usually incorrectly called *target identification*). The fighter and wingman will coordinate modes through the net so that both have situational awareness during the long time span required to provide target recognition.

Another important fighter category is weapon support. Missile update is the measurement of missile and target position, velocity and acceleration to allow statistically independent measurements for transfer alignment, as well as missile state-of-health. Missile update provides the latest target information and future dynamics prediction by data link. IR missile slaving co-aligns radar and seeker. Since gun effective ranges are very short, gun ranging causes the radar to sense the gun field of fire, predicts angle rate, and measures range to a target for tentative gunfire.[9] It may also track gun rounds during fire.

There are thousands of electrical degrees of phase between free space and the A/D converters. The combination of temperature, time, and manufacturing tolerances gives rise to the need for self calibration, test, fault detection, failure diagnosis, and needed corrections, which are performed by a subsuite of performance monitor software.

Timing Structure. The significance of the remaining parameters in Tables 5.1 and 5.2 can best be illustrated with a timing structure typical of fighter radars.[7,8,9] Figure 5.14 shows a modern radar timing structure in a sequence of progressively expanded timelines. The first row of Figure 5.14 shows a typical scan cycle covering the required volume of interest for a specific mode. The time span for a full scan cycle might be 1 to 5 seconds. Inside the total scan cycle time, there may be several bars of a scanned region of space with a time span of a few tenths of a second. A bar is a scan segment along a single angular trajectory, as shown in Figure 5.20, later in the chapter.

TABLE 5.2 Typical Waveform Parameters A-A Modes[9]

Radar Modes	PRF (kHz)	Pulse Width (μsec)	Duty Ratio (%)	Pulse Comp. Ratio	Freq. Looks	Pulses Per CPI	Instant Band-Width (MHz)	Total Pulses in T_{OT}
Range-Gated High PRF	100–300	1–3	10–30	1–13	1–4	500–2000	0.3–10	1500–6000
Medium PRF	6–20	1–20	1–25	5–256	1–4	30–256	1–10	250–2000
Burst Ranging	3–20	2–60	0.1–25	1–256	1–4	20–256	0.5–15	20–550
Active Track	8–300	0.1–20	0.1–25	1–256	4–8	1–64	1–50	20–1000
Raid Assessment	2–16	2–60	0.1–25	1–256	1–4	20–256	0.5–15	20–1000
Non Coop. Target Rec.	2–20	1–200	0.1–10	1–16384	1–4	20–256	0.2–100	20–1000
HiPwr Jam	50–300	3–10	10–50	13–512	1–8	1–8	1–100	200–2k
Cal/AGC	2–300	0.1–60	0.01–50	1–16384	1–8	8–64	0.2–500	8–64
Air Data Link	10–300	1–20	1–33	1–16	1	100–500	0.1–1	100–500
Gun Ranging	10–20	0.1–0.5	0.1–1	1–5	1–4	4–32	1–10	4–128
Weather Avoidance	0.5–5	1–50	1–10	1–13	1–2	1–8	0.1–1	1–16

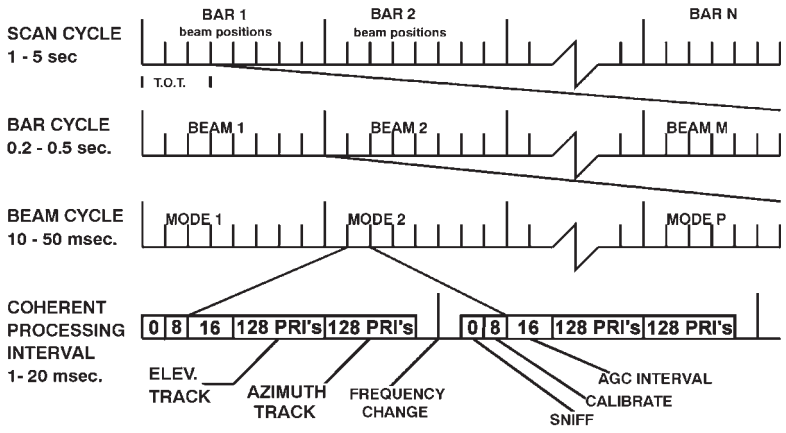


FIGURE 5.14 Typical MFAR timing sequences[7,8,9] (*Courtesy SciTech Publishing*)

Each bar consists of multiple beam positions of a few tens of milliseconds each, which are computed on the fly to optimally cover the selected volume. Each beam cycle, in turn, may contain one or more radar modes or submodes, such as those contained in Tables 5.1 or 5.2 and depicted in the lowest line of Figure 5.6. The modes may not be invoked each time depending on the geometry between the aircraft and the intended target set.

The mode time is broken up into coherent processing intervals (CPI's). A coherent processing interval is segmented, as shown in the bottom row of Figure 5.14. The particular example shown is tracking that might be used in FTT, GMTT, PVU, or A-G Ranging, as shown previously in Figure 5.10, and later in Figures 5.32 and 5.38. It consists of a frequency change; settling time; passive receiving to be sure the band isn't jammed; calibrate that doesn't intentionally radiate but often there is some RF leakage radiated; an automatic gain control (AGC) interval in which a number of pulses are transmitted to set the receiver gain; and finally two intervals in which range, doppler, and angle discriminants are formed. These CPIs often but not always have constant power, frequency sequence, PRF sequence, pulsewidth, pulse compression, and bandwidth.[7,8,9]

5.3 A-A MODE DESCRIPTIONS & WAVEFORMS

Air-to-Air Search, Acquisition and Track— Medium PRF. It may be instructive to examine how several modes are generated and processed to understand why the waveforms must be the way they are. Medium PRF trades long-range detection performance (see Figure 5.21, later in the chapter) for all aspect target detection.[28,52,53] Often high and medium PRF waveforms are interleaved on alternate scans (see Figure 5.20) to improve total performance.[28,54,55] After 30 years of searching for an optimum set, most modern medium PRF modes have devolved to a range of PRFs between 8 and 20 kHz in a detection set of 8 for the time on target.[44,56–61] These PRFs are chosen to minimize range and velocity blind zones while simultaneously allowing unambiguous resolution of target range and doppler returns in a sparse target space.[62,63,64] Range blind zones are those ranges in which a target is eclipsed by the transmitted pulse. Velocity or doppler blind zones are those velocities or dopplers that are excluded due to the

main-beam clutter and ground moving target filter rejection notch. Target detection requires detections in at least 3 of the 8 PRFs with all PRFs clear at maximum range. The PRF selection criteria usually requires that the PRF set is 96% clear—in other words, at least a specified number (typically 3) of PRFs must have an above threshold return echo for the minimum specified target for the full specified range-doppler coverage.

A typical processing block diagram is given in Figure 5.15. Each PRF processing interval is different, but they average out to an optimum, as shown later in Figure 5.17. Both main and guard channel processing is required to reject false targets.[25] Some STAP processing may have been performed before this process, but traditional sidelobe and main-beam clutter is less of a limit than ground moving targets, which have very large cross sections and exo-dopplers (i.e., doppler far enough out of main-beam clutter that detection is not limited by the clutter return). MPRF usually has a small amount of pulse compression (1:1 to 169:1), which still may require doppler compensation.[65] Main and guard channels are processed in the same way. Obviously, the two spectra are quite different and separate false alarm and noise ensemble estimates are made. This leads to separate threshold settings. Multiple channels are used to estimate interference and select ECCM strategy. Main channel detections are examined for GMTs and centroided in range and doppler (because a return in range or doppler may straddle multiple bins, the centroid of those returns in multiple bins must be estimated from the amplitude in each bin and the number of bins straddled). The guard channel is detected and the thresholded results are used to gate the main channel results for the final hit-miss count. Genuine targets are resolved in range and doppler, presented to a display and used for TWS correlation and tracking.[8]

False alarms are a critical issue in most radar modes. These are usually suppressed for thermal noise by constant false alarm rate thresholding, coincidence detection, and post-detection integration with frequency agility. Clutter false alarms are suppressed by adaptive aperture tapering, low-noise front-end hardware, wide dynamic range A/Ds, clutter rejection filtering (including STAP), pulse compression sidelobe suppression, doppler filter sidelobe control, guard channel processing, radome reflection lobe compensation, angle ratio tests (see Figure 5.37 and the "fringe region" for an example angle-ratio- test), and adaptive PRF selection.

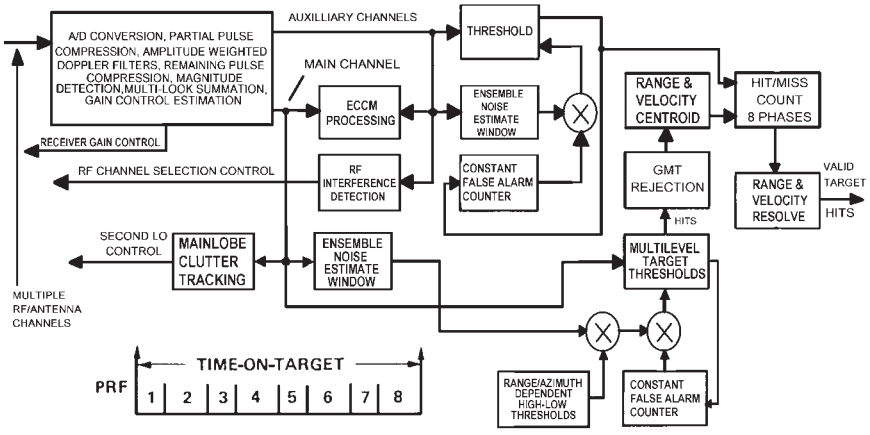


FIGURE 5.15 Typical MPRF processing (*adapted[8]; courtesy SciTech Publishing*)

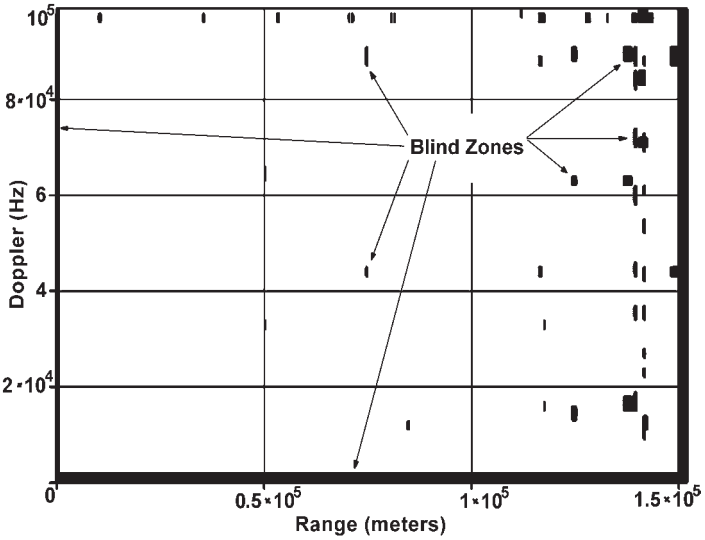


FIGURE 5.16 Medium PRF range-velocity blind zones

MPRF-Typical Range-Doppler Blind Map. For example, a typical MPRF set for X band with range-doppler coverage of 150 km–100 kHz is shown in Figure 5.16. This set is for a 3° antenna beamwidth, ownship (i.e., the radar carrying fighter) velocity of 300 m/s, and an angle off the velocity vector of 30°. The PRF set is 8.88, 10.85, 12.04, 12.82, 14.11, 14.80, 15.98, and 16.77 kHz. Historically, a PRF set was calculated during design and remained fixed during deployment. Modern multifunctional radar computing is so robust that PRF sets can be selected in real time based on situation geometry and look angle. The set, which generated Figure 5.16, on the average is clear on 5.6 out of 8 PRFs for a single target. Except for two small doppler regions, all the PRFs are clear at maximum range, which provides maximum detection and minimum loss at the design range. For some pulse compression waveforms, the eclipsing loss is almost linear and partial overlap still allows shorter-range detection. Eclipsing loss is that diminishment of received power when the receiver is off during the transmitted pulse. It is often the largest single loss in high duty ratio waveforms. The bad news is that the average detection power loss is slightly over 3 dB (see Figure 5.21).

MPRF Selection Algorithms. Obviously, selecting PRFs in real time requires several rules to get close to a final set. This is followed by small iterations to pick the optimum set. For medium PRF, both range and velocity blind zones are important.[52,65] First, the software must pick a central PRF about which all the other PRFs are deviations to fill out the desired visibility criteria. Second, the PRF set should all be clear at the maximum design range so that detection losses are at a minimum.

Figure 5.17 shows one example criteria for selecting the central PRF, i.e., the highest probability of visibility (P_V).[45] In the example, the product (P_V) of the range (P_R) and doppler (P_D) target visibility probabilities for a single PRF peaks at approximately 0.47 and thus the other PRFs must fill in to reach 96% clear or higher. There are several

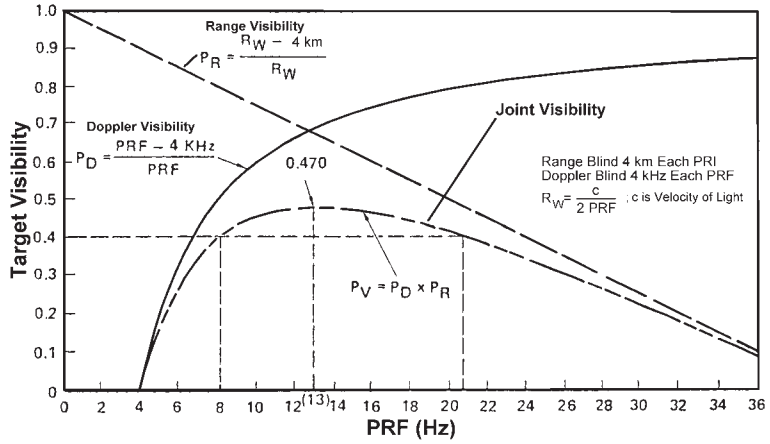

Range Visibility
$P_R = \frac{R_W - 4\text{ km}}{R_W}$
Doppler Visibility
$P_D = \frac{PRF - 4\text{ KHz}}{PRF}$
Joint Visibility
0.470
$P_V = P_D \times P_R$
Range Blind 4 km Each PRI
Doppler Blind 4 kHz Each PRF
$R_W = \frac{c}{2\,PRF}$; c is Velocity of Light
Target Visibility
PRF (Hz)

FIGURE 5.17 Medium-PRF central PRF selection example[45]

other factors to be considered: doppler and range blind zones and eclipsing and sidelobe clutter. Even with STAP, sidelobe clutter is a major limitation.[67,68,93] Both sidelobe and main-beam clutter can be minimized by narrow doppler and/or range bins (i.e., resolution cells), which imply longer dwell times and higher transmit bandwidth.

One example method for selecting a set of PRFs for MPRF is given in Eq. 5.1. The basic idea is to find a time interval, T_A, representing the desired maximum clear range, and then to choose a set of PRIs in which all will be clear at maximum range. This can be achieved by dividing T_A by an integer, typically 9 to 17. This set will generally not provide 96% clear over the range-doppler space. The even divisor PRIs can be perturbed iteratively by a small amount to achieve the desired visibility. The normalized target signal-to-noise ratio, TP, varies dramatically with straddle and eclipsing losses (for example, see Figure 5.18). The function to be optimized is a thresholded version of $TP_{k \text{ or } j}$.

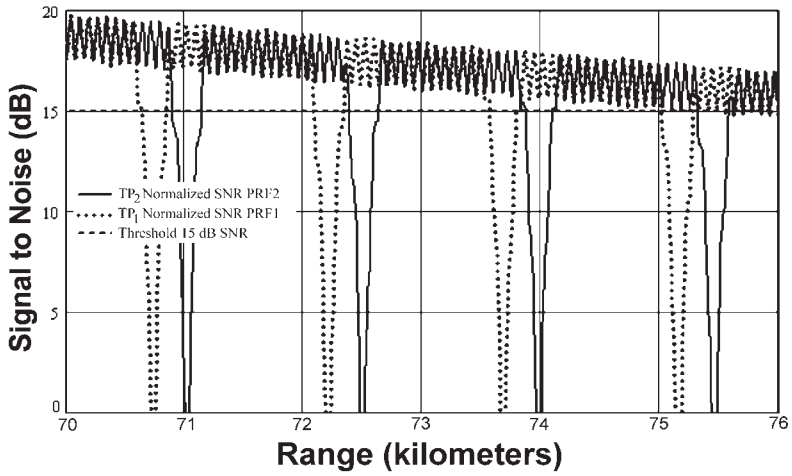


FIGURE 5.18 Example RGHPRF eclipsing and straddle near maximum range

For example, the threshold scheme might be 15 dB SNR per PRI and 3 out of 8 for all PRIs. Often multiple and different thresholds are used for each CPI and PRI. Lower thresholds are allowable for higher numbers of total hits.[28] It should be noted that eclipsing and straddling, and so on, have much less effect at closer ranges where there is usually more than enough SNR. Another serendipitous effect of this selection technique is that as an individual PRI range clear region gets smaller, the doppler clear region gets larger, filling-in the blind zones in both dimensions.

$$T_A = 2 \times \left(\frac{R_c}{c} + \tau_p \right), PRI_k = \frac{T_A}{C_1 + 2 \times k}, PRI_j = \frac{T_A}{C_2 + 2 \times j + \delta_j}$$
$$TP_{k \text{ or } j}(f,r) = \frac{C_3}{r^4} V_{\text{blind}}[\text{mod}(f, 1/PRI_{k \text{ or } j})] \times R_{\text{blind}}[\text{mod}(r, PRI_{k \text{ or } j})] \quad (5.1)$$

where: R_c is maximum design clear range
c is the velocity of light $2.9979 \cdot 10^8$ m/s
τ_p is transmitted pulse width, k and j are indices e.g. 0…4
C_1 is an odd integer e.g. 9, C_2 is an even integer e.g. 12
δ_j is a small perturbation e.g. 0.1…0.3 yielding visibility > 96%
V_{blind} is a function of f describing eclipsing and straddling
R_{blind} is a function of r describing eclipsing and straddling
C_3 is a constant representing the remainder of the range equation
f is frequency, r is range, mod is modulo the first variable by the second

Range-Gated High PRF. Range-gated high PRF (RGHPRF) performance is dramatically better for detection of higher speed closing targets.[44,54,55,70] (Range gates are often smaller than range resolution cells or bins). RGHPRF produces the longest detection range against closing low cross section targets.[71] Ultra-low noise frequency references are required to improve subclutter visibility on low RCS targets even using STAP. Range gating dramatically improves sidelobe clutter rejection, which allows operation at lower ownship altitudes. Principal limitations of RGHPRF closing target detection performance are eclipsing (a radar return when the receiver is off during the transmitted pulse) and range gate straddle losses (the range gate sampling time misses the peak of the radar return).[15] Figure 5.18 shows TP_i with eclipsing and straddle losses near maximum range for a high performance RGHPRF. This mode is optimized for low cross section targets out to just beyond 75 km maximum range. The particular example has overlapping range gates to minimize straddle loss and two PRFs to allow at least one clear PRF near maximum range. The PRFs are 101.7 kHz and 101.3 kHz. Duty ratio is 10% with 15 dB required detection SNR. Averaged over all possible target positions and closing dopplers, the losses for this mode are a surprisingly small 0.4 dB.

The range-doppler blind zones plot is shown in Figure 5.19 corresponding to the Figure 5.18 waveform. Compared to the medium PRF plot shown in Figure 5.16, the clear region (and corresponding losses) is dramatically better. Unfortunately, range is very ambiguous. Normally, a RGHPRF range-while-search (RWS) mode is interleaved with the highest performance velocity-search (VS) mode to range on previously detected targets.

Often, RWS is RGHPRF with three phases in which a constant frequency and two chirp (linear FM) frequencies (triangular up-down or up-steeper up) are used to resolve range and doppler in a sparse target space. At low altitudes, sidelobe clutter, even with STAP processing, limits performance for all targets but especially opening targets.

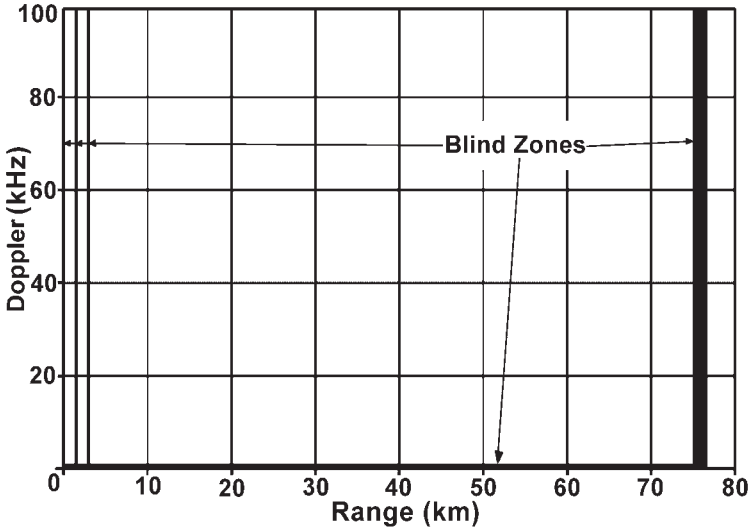


FIGURE 5.19 RGHPRF range-velocity blind zones, corresponding to Figure 5.18 waveforms

This limitation leads to the need for another mode interleaved with RGHPRF. Fortunately, the timeline for opening targets is much longer (net speed is less) and the engagement range is much shorter (weapon closure rates are too slow).

Often in general search, MPRF-VRS (medium PRF velocity-range search) is interleaved with HPRF VS and RWS, as shown in Figure 5.20, to provide all aspect detection. Unfortunately, both RWS and VRS have poorer maximum detection range. RGHPRF can provide all aspect detection but tail performance is dramatically poorer due to sidelobe clutter. Even with STAP, which significantly improves sidelobe clutter rejection, low altitude tail aspect detection for RGHPRF is poorer.[44,45,55]

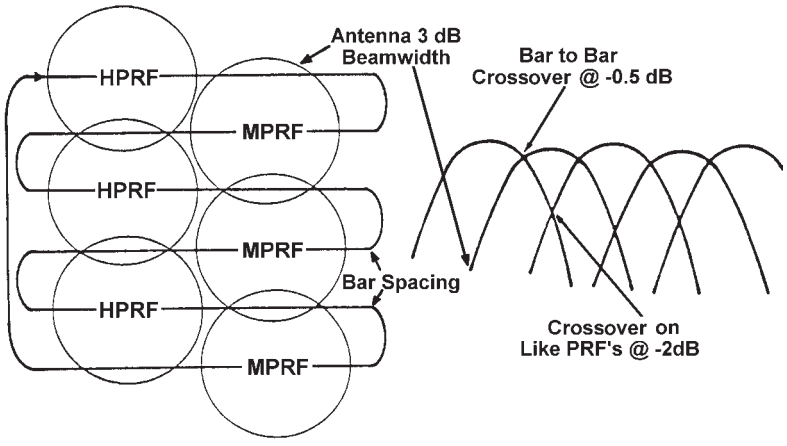


FIGURE 5.20 High and medium PRF interleave for all aspect detection[45]

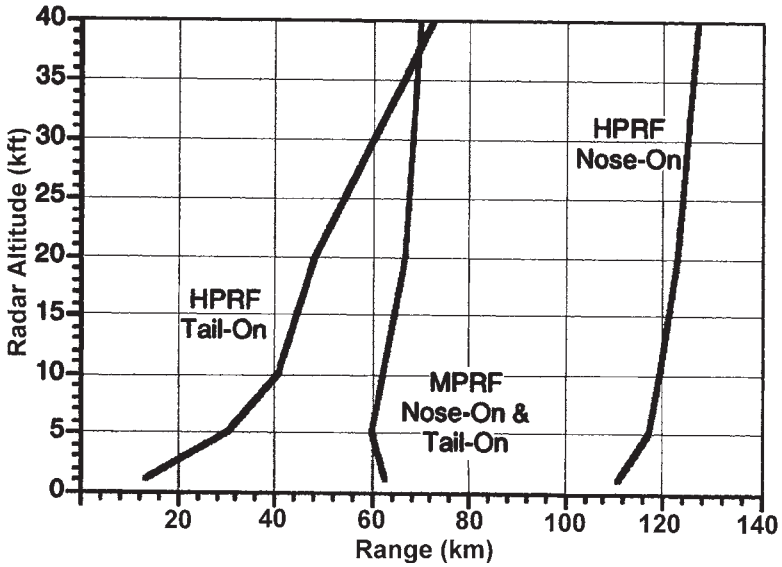


FIGURE 5.21 Comparison of high and medium PRF[9]

An example comparison of HPRF and MPRF as a function of altitude for a given maximum transmitter power, power-aperture product, and typical antenna and radome integrated sidelobe ratio is shown in Figure 5.21. At high altitude and nose-on, there is more than an 11 dB difference caused by blind zones, straddle, folded clutter, processing, and thresholding losses.[9,11,28]

RGHPRF Selection Algorithms. First, as in the MPRF case, all PRFs should be clear at the maximum design range. Second, all PRFs should be clear to the maximum doppler of interest. One possible selection criteria is given in Eq. 5.2. Although the details are quite different, the basic philosophy in PRF selection is to optimize long-range clear regions.

$$T_A = \frac{2 \times R_c}{c} + \tau_p \text{ and } PRI_A = \frac{0.25 \times \lambda}{V_a + V_t} \text{ and } I = \text{ceil}\left[\frac{T_A}{PRI_A}\right]$$

$$\text{then } PRI_1 = \frac{T_A}{I} \text{ and } PRI_2 = PRI_1 \times \left[\frac{c \times \tau_p}{R_c} + 1\right] \tag{5.2}$$

where: R_c is maximum design clear range,
c is the velocity of light 2.9979×10^8 m/s,
τ_p is transmitted pulse width, λ is transmitted wavelength,
ceil is the next integer above the value of the variable,
V_a and V_t are the maximum velocities of interest for aircraft and target respectively.

Noncooperative Air Target Recognition[72] MFAR target recognition (TID) recognizes target type but not unique identification. There are cooperative target

identification methods such as JTIDS, IFF, and RF tagging that can be unique. TID depends on detecting features of the radar signature in fusion with emissions and other sensors. The five most common TID signatures are monopulse extent (similar to the example shown in Figure 5.25), resonances, high resolution range (HRR) profiles, doppler spread, stepped frequency waveform modulation or multifrequency (SFWM/MFR), which can be transformed into a range profile, and inverse synthetic aperture radar (ISAR).[16,45] Monopulse extent allows estimation of length and width as well as separation of closely spaced aircraft. A high range resolution profile also allows the separation of targets flying in close formation as well as the separation of a missile from a target. A high range resolution profile on a single target can allow recognition, assuming the target attitude is known or has been guessed. Length, width, and location of major scattering features can be projected into a range profile if the attitude is known. The number of types of major civilian and military aircraft and ships is at most a few thousand, easily storable in memory. Unfortunately, recognition is limited to broad categories rather than MIG-29M2 versus MIG-29S (even though there are significant differences that air show visitors can easily see[89]).

The basic notion of doppler, resonances, stepped (SFWM), and multifrequency (MFR) signatures is modulation either by reflections from moving parts, e.g., engine compressor, turbine, rotor, or propeller blades, or by interactions from scatterers along the aircraft or vehicle, e.g., fuselage, wing, antennas, or stores. SFWM/MFR signatures are closely related to high range resolution signatures (a Fourier transform easily converts one to the other), and they suffer the same attitude estimation limitations. The principal advantage to MFR is that many deployed radars have multiple channels and switching between them on a single target is relatively easy. A simplified version of the recognition process is summarized in Figure 5.22.

Doppler signatures require high doppler resolution, which is usually easily achieved and limited only by dwell time. The individual scatterers, which give rise to doppler spread, are small and so recognition is usually limited to a fraction (1/2 typical) of maximum range. Jet engine modulation (JEM), a subset of doppler signatures, is an excellent target recognition method. Even aircraft, which use the same engine type, often have variations in the engine application, such as the number of compressor blades or number of engines, which allows unique type recognition. The real picture of JEM is not so clean because of multiple on-aircraft bounces, straddling, and speed variations, but centroiding of each line improves the signature estimate. The last method of TID, ISAR, will be dealt with in another section. ISAR works well on both aircraft and ships. A typical tail hemisphere air-to-air ISAR is shown in Figure 5.23.

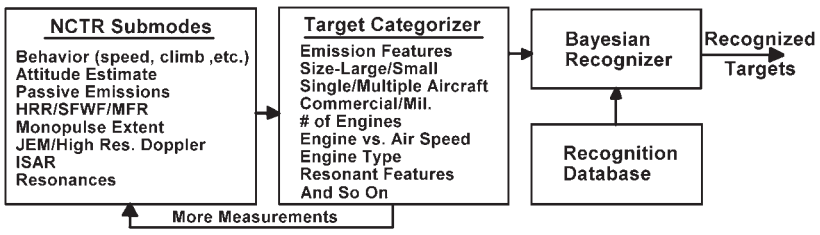


FIGURE 5.22 Noncooperative target recognition submodes[45]

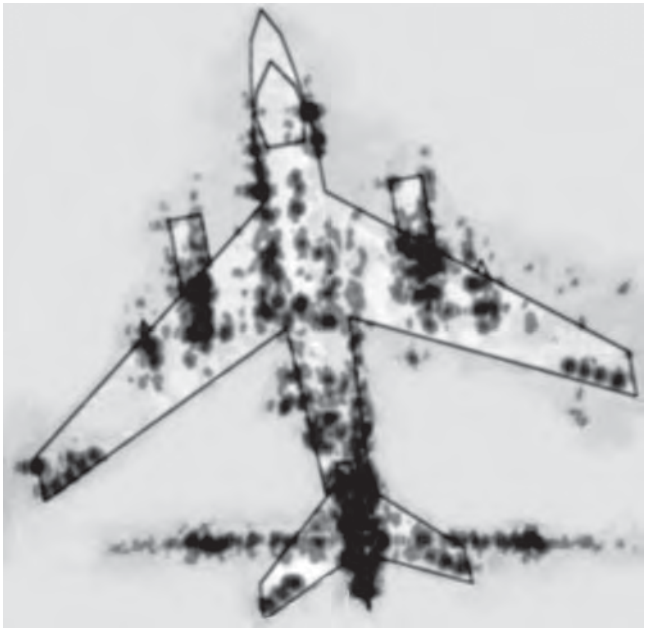


FIGURE 5.23 A-A ISAR example TA-3B[45]

The fusion of the recognition of each of the signatures above provides excellent noncooperative recognition.

Weather Avoidance. Many aircraft have separate weather radars. Weather avoidance is normally incorporated into modern fighter radars. The normal operating frequency for a fighter radar has not been considered optimum for weather detection and avoidance[15]—primarily due to lack of penetration depth into a storm and reduced operating range. However, with complex atmospheric attenuation compensation and doppler methods, weather can be detected well enough to allow warning and avoidance of storms. The principal challenge is compensating for backscatter from the leading edge of a storm and adjusting for attenuation to see far enough into a storm to evaluate its severity. The backscatter from each cell is measured, the power remaining is calculated, the attenuation in the next cell is estimated, and then the backscatter in the next cell is measured, and so on. When the power in the cells drops to the noise level, those cells behind it are declared blind. Since penetration range into a storm is not great, the MFAR weather mode usually has provisions to mark the last visible or reliable range on the weather display. This is so the pilot does not fly into a dark area believing there is no weather.

Air Data Links. The MFAR is part of a network of sensors and information sources (C3ISR net) sometimes called the *Global Information Grid (GIG)*. A major use of radar and aircraft data links is to provide total situational awareness.[21] By using on-board and off-board sensor fusion, a total air and ground picture can be presented in the cockpit. This picture can be a combination of data from

other radar sensors (wingman or support aircraft) on similar platforms to reports by observers with binoculars. Because the modern fighter is net-centric, using everything available on- and off-board the aircraft, net-centric operation requires dramatically higher levels of data exchange and fusion of data for presentation to the operator. Radar modes can be scheduled between multiple aircraft in real time through the data links.

The two main uses for data links associated with high performance aircraft are high bandwidth imagery transmission from a weapon or sensor platform to a second platform or ground station and low bandwidth transmission of context, targeting data, guidance, and housekeeping commands.[74–79] The largest quantity of data links are associated with weapons. The waveform selected to transmit this and other data must not compromise the signature of the platform at either end of the link.[8,34–37]

There are numerous data links on fighters. Table 5.3 shows air data links that might be on a fighter platform. In spite of this fact, the radar or part of its aperture is often used for a data link, especially to missiles on the fly and in response to peacetime air traffic control interrogations. Pulse amplitude (including on-off), pulse position, phase shift, and frequency shift modulation are commonly used. Links may be unidirectional or bidirectional. Some missiles require semi-active illumination as well as reference signals and midcourse command data derived from missile and target tracking. The data to and from the missile is often an encrypted phase-code in or near the radar operating band. In some cases, the frequency channel is randomly selected at the factory and hardwired into the missile. Frequency channels are typically selected or communicated to the radar immediately before launch. If the data link frequency is well below the radar band, usually a small number of radiators at that lower frequency are imbedded in the radar aperture. If the frequency is close enough to the radar band, the radar aperture or a segment of the aperture is used.

Radar Aperture Datalinking.[81] Historically, datalink functions embedded in MFARs have been used for the midcourse guidance of missiles. An emerging application is the use of the radar aperture as a high power, high gain primary datalink antenna,

TABLE 5.3 Air Data Links[80]

Link	Freq Band	Data Rate (kb/s)	ECCM
ARC-164	UHF	1.8	High
ARC-126	UHF	1.8	High
ARC-201	VHF	8	Moderate
ARC-210	VHF/UHF	8	Moderate–High
TADIL	UHF-L	1.8–56.6	Moderate–High
JTIDS	L	28.8–56.6	Moderate
JTIDS LET	L	383.6–1,180.8	Moderate
JTRS	VHF-X	1.8–1,544	Moderate–High
TADIXS	UHF	9.6	Moderate
MFAR	X-Ku	2–105	Moderate–High
Milstar	UHF, Ku - Ka	4.8–1,544	High
TCDL	X - Ku	1,000–256,000	Moderate

where datalink transmission and reception are interleaved with other modes. The principal limitation of most general-purpose datalink equipment is the low power-aperture performance associated with omnidirectional, often shared, antenna apertures and limited power levels. This constrains achievable data transfer rates, regardless of channel bandwidth. An associated problem is vulnerability to intercept and jamming inherent in widebeam apertures. An X or Ku band MFAR can emit power levels in the multi-kilowatt range with main-beam beamwidths of a few degrees, affording high data rates and significant resistance to jamming and intercept. Transmit data rates of over 500 Mbps, and receive data rates of up to 1 Gbps have been demonstrated using a production AESA and a modified Common Data Link (CDL) waveform.[82] Modeling using representative MFAR parameters indicates that performance bounds are at several Gbps throughput over distances in excess of 400 nautical miles, subject to MFAR performance, platform altitude, tropospheric conditions, and forward scattering effects.[81]

Implementation requires accurate antenna pointing, since there is relative motion with respect to the other end of the link. One technique involves the use of an out-of-band datalink channel, e.g., JTIDS, to carry GPS position updates.[83] Doppler shifting due to link geometry dynamics must be actively compensated. A related issue is synchronization in time to allocate transmission and reception windows and to synchronize timebases. When existing waveforms must be used, this can present challenges. Existing aperture scheduling algorithms can then allocate time for transmission or reception.[81]

To achieve very high throughputs, phase linearity in transmit and receive paths is critical since data transmission waveforms rely on modulation that is every bit as complex as many radar modes. This can also impact choice of taper function because angular variations in phase across the main-beam wavefront may incur performance penalties. Where the MFAR is phase steered, aperture fill and sidelobe steering effects constrain usable aperture bandwidth similar to SAR limitations. The latter is because the element phase angles required to point the main beam are not the same as those for the outer sidelobes of the modulation used.[81]

Low bandwidth data links can use all the radar bandwidth to improve encryption and signal-to-jam ratios. However, the data link on a weapon is traveling to the target, which will inevitably attempt to protect itself. When the weapon is near the target, the signal-to-jam ratio can be very unfavorable. Antenna jammer nulling is usually required since transmitting more power to burn through may not be possible. Clearly, the data from and to a weapon must also be sufficiently encrypted to prevent take-over of the weapon in flight.[8]

Time synchronized with a radar transmission on a different set of beams and/or frequencies, messages are sent to one or more missiles on the fly to the targets. Obviously, all the random frequency diversity, spread spectrum, and encryption necessary for robust communication should be incorporated into the message.[78] Each missile may answer back at a known but randomized offset frequency and time with image or housekeeping data. Again a waveform as robust as possible is used, but since the baseband data and link geometry may be quite different, the data compression, diversity, and encryption may be different.

The missile datalink waveform usually must be stealthy and greatly attenuated in the direction of the target since one countermeasures strategy is a deception repeater jammer at the target. High accuracy time and frequency synchronization, including range opening and doppler effects between both ends of the link can dramatically reduce the effectiveness of jamming by narrowing the susceptibility window. Time and frequency synchronization also minimizes acquisition or reacquisition time.

An aircraft using a data link is moving with respect to the other end of the link, so the link geometry is continually changing in time, frequency, aspect, and attitude. The signal processor will generate waveforms for transmission by the seeker or data link. It will also measure target range, angle, doppler, and so on, and provide those to the other platform. The MFAR signal processor sends motion sensing and navigation estimates to correct measurements; to track, encode, and decode datalink messages; and to perform jammer nulling.

Beacon Rendezvous and Station Keeping. Most modern military aircraft depend on in-flight refueling for many missions. This requires rendezvous with tanker aircraft during all weather conditions as well as station keeping until aircraft currently in line for refueling depart. This may involve detecting a coded beacon on the tanker, skin tracking tankers and other aircraft at close range. Station keeping ranges can be between 30 and 1000s of meters. Special short-range radar modes are usually used for this purpose. Low power, short pulse or FM-CW waveforms are often used. One meter accuracy and 30 meter minimum range is usually required for blind tanking.

High Power-Aperture Jamming. The basic notion behind MFAR high power-aperture jamming is suggested in Figure 5.24.[9,11,71,84]

A threat emitter, whether surface or airborne, is first detected and recognized by the spherical coverage radar warning receiver (RWR) function (possibly just an application overlay on the RF and processing infrastructure shown in Figure 5.4). If the intercept is inside the radar field of view (FOV), fine angle-of-arrival (AOA) and possibly burst ranging are performed with the primary radar aperture, as shown in the top portion of Figure 5.24. High-gain electronic support measures (ESM) are then performed and recorded on the emitter main beam or sidelobes using the nose aperture. If it is determined from an on-board threat table, current rules of engagement, or mission plan, high power density jamming based on the corresponding on-board techniques table may be initiated using the high-gain nose aperture.[84]

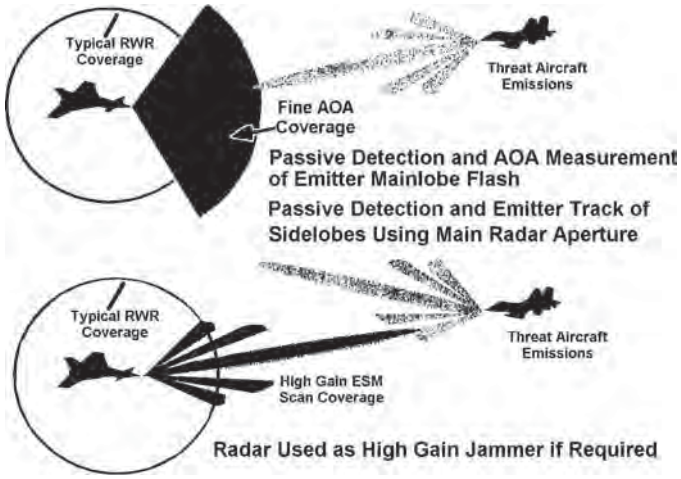

Typical RWR Coverage
Threat Aircraft Emissions
Fine AOA Coverage
Passive Detection and AOA Measurement of Emitter Mainlobe Flash
Passive Detection and Emitter Track of Sidelobes Using Main Radar Aperture
Typical RWR Coverage
Threat Aircraft Emissions
High Gain ESM Scan Coverage
Radar Used as High Gain Jammer if Required

FIGURE 5.24 MFAR ECM example[9]

Because the adversary radar may also be a MFAR, threat tables will be required to categorize them by their apparent statistical nature. Old style matching by PRF, pulsewidth, and pulse train envelope won't work very well because waveforms vary so much. The typical nose aperture radar–based effective radiated peak power (ERPP) can easily exceed 75 dBW, which is normally more than enough to play hob with threat radars.[39,85] For example, assuming a 10 GHz in-band signal, –3 dBi threat sidelobe, and –110 dBW threat sensitivity, a jamming pulse 60 dB above minimum sensitivity can be generated at 20 km. Obviously, in the near sidelobes or main beam, the range for a 60 dB pulse will be much greater.[84]

5.4 A-S MODE DESCRIPTIONS & WAVEFORMS

Terrain Following, Terrain Avoidance. The next example is terrain following, terrain avoidance (TF/TA) shown in Figure 5.25. In terrain following (TF), the antenna scans several vertical bars oriented along the aircraft velocity vector and generates an altitude-range profile that is sometimes displayed to the pilot on an E-scan display. Depending on the aircraft's maneuvering capabilities, there is a control profile 1g acceleration maneuver control line shown as an upward curving line in the upper right of Figure 5.25.[86–88] If this conceptual line intercepts the terrain anywhere in range, an automatic up maneuver is performed. There is also a conceptual pushover line, not shown in the figure, which causes a corresponding down maneuver. The control profile in modern aircraft is automatic because a human pilot does not have the reflexes to avoid all possible detected obstacles.

In terrain avoidance (TA), the antenna scan is in a horizontal plane (shown in the upper left of Figure 5.25). Several altitude plane cuts are estimated and presented to

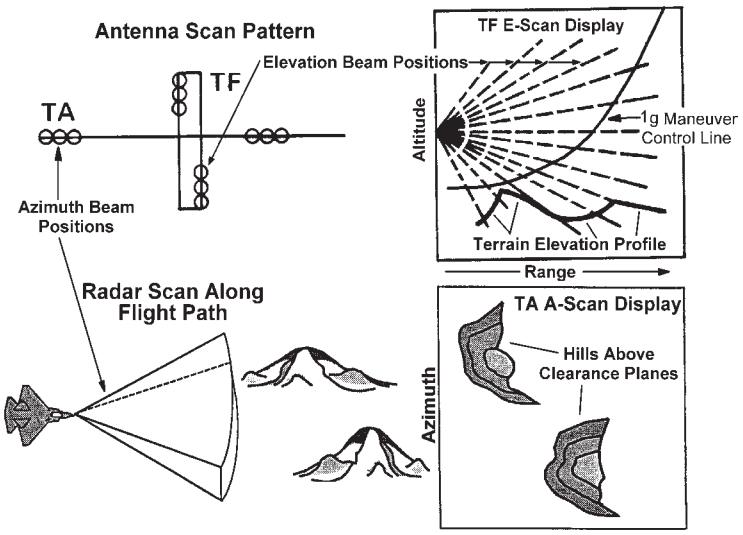


FIGURE 5.25 TF/TA mode example (*adapted*[8]*; courtesy SciTech Publishing*)

the pilot on an azimuth-range display (shown in the lower right of Figure 5.25). The terrain avoidance scan pattern shows all the terrain that is near or above the flight altitude and one cut below at a set clearance altitude (500 ft. typically).[86] Figure 5.25 (lower left and lower right) shows the situation geometry of an aircraft flying toward two hills and the corresponding altitude cuts displayed to the pilot. This allows either manual or automatic turning flight to maintain a lower altitude.

TF/TA allows an aircraft to penetrate at low altitude using the terrain as masking, thus preventing early detection. TF/TA is an important aspect of stealth even when the altitude is not all that low because lower altitudes provide some terrain obscuration and many other competing targets with similar cross-sections.[8]

Terrain Height Estimation. Some of the features of TF/TA are the required scan pattern, the number of independent frequency looks required to obtain a valid estimate of the height of a possibly scintillating object along the flight path, and the range coverage. Because terrain height is estimated through an elevation measurement, angle accuracy is critical. The range coverage, although short, requires multiple overlapping beams and multiple waveforms. One method for calculating terrain height[8] is shown in Figure 5.26. It consists of measuring the centroid and extent of each individual beam position over many pulses and estimating the top of the terrain in each beam, as shown in the figure. The calculation is summarized in Eq. 5.3.

$$P_r = \sum_i |S_i|^2 \text{ power received, } C_r = \text{Re}\left\{\frac{\sum_i S_i \times D_i}{P_r}\right\} \text{centroid}$$
$$E_r^2 = \frac{\sum_i |D_i|^2}{P_r} - C_r^2 \text{ extent squared, } T = C_r + 0.5 \times E_r \text{ terrain top estimate} \tag{5.3}$$

where S_i is a single sum monopulse measurement, D_i is the corresponding elevation difference monopulse measurement.

Usually the range-elevation profile is measured in multiple segments with separate PRFs and pulsewidths. The lowest PRF is used to measure the longest-range portion of the profile at the top of the elevation scan. It uses the largest pulse compression ratio (16:1–32:1). Each beam position overlaps by as much as 90% and multiple frequency

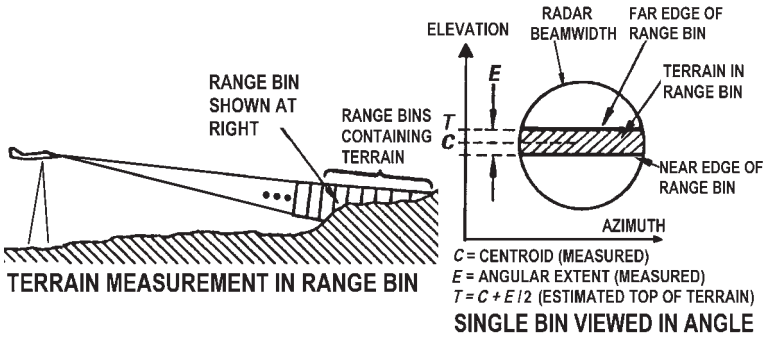


FIGURE 5.26 Terrain height estimation[8] (*Courtesy SciTech Publishing*)

looks in each beam create as many as 64 independent looks. The shortest range at the bottom of the elevation scan uses a short pulse with no pulse compression and a much higher PRF, but the same number of looks. The pulses in a T_{OT} are all the pulses that illuminate a single spot from the overlapping beams. Each overlapping beam must be compensated for the antenna look angle before the beams can be summed for a terrain height estimate from all the beams.[8] The radar cross section of the terrain could be quite low (e.g., snow-covered level treeless terrain), so some pulses may be integrated coherently to improve signal-to-noise ratio for a CPI of up to 8 pulses, as shown in the TF/TA entry in Table 5.1.

Terrain Database Merging. For the purposes of safety as well as stealth, active radar measurements are merged with a prestored terrain database.[87] Figure 5.27 shows the general concept of merged TF/TA measurements with stored data.

Active radar measurements are made out to a few miles. The instant use terrain database extends out to perhaps ten miles. The terrain database cannot be completely current and may contain certain systematic errors. For example, the database cannot contain the height of wires strung between towers or structures erected since the database was prepared. For the lowest possible flight profiles with less than 10^{-6} probability of crash per mission, the prestored data is merged and verified with active radar measurements. Low crash probabilities may also require some hardware and software redundancy. In addition, as the aircraft flies directly over a piece of terrain, combined terrain profile is verified by a radar altimeter function (TERCOM/TERPROM) in the RF and processor complex. Usually the prestored data is generated at the required resolution before a mission from the worldwide digital terrain elevation database (DTED).

Sea Surface Search, Acquisition, and Track. Sea surface search, acquisition, and track are oriented toward three types of targets: surface ships, submarines snorkeling or near the surface, and search and rescue. Tracking may be preliminary to attack with antiship weapons. Although most ships are large radar targets, they move relatively slowly compared to land vehicles and aircraft. In addition, sea clutter exhibits both current and wind-driven motion as well as "spiky" behavior. These facts often require high resolution and multiple looks in frequency or time to allow smoothing of sea clutter for stable detection and track.[16,45] If the target is a significant surface vessel, then RCS might be 1000 m^2, and a 30 m range resolution might be used for search

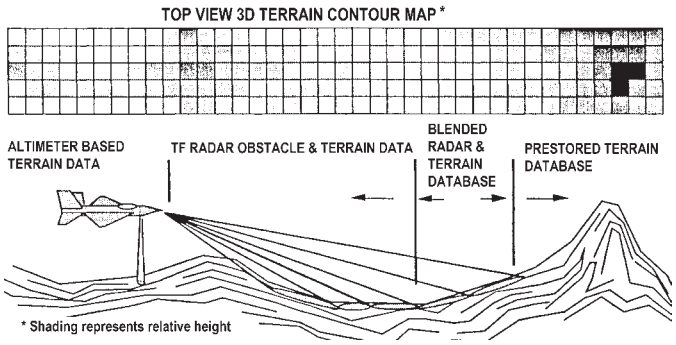

TOP VIEW 3D TERRAIN CONTOUR MAP *
ALTIMETER BASED TERRAIN DATA
TF RADAR OBSTACLE & TERRAIN DATA
BLENDED RADAR & TERRAIN DATABASE
PRESTORED TERRAIN DATABASE
* Shading represents relative height

FIGURE 5.27 TF/TA terrain merging[7,8,87] (*Courtesy SciTech Publishing*)

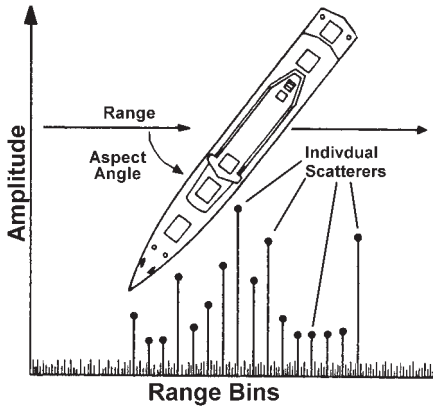


FIGURE 5.28 Range profile ship recognition[45]

and acquisition. If the target is a periscope or person in a life raft then 0.3 m resolution might be used since the RCS might be less than 1 m^2 and smoothing is especially important. DPCA and doppler processing is often interleaved with traditional bright (20 dB or greater above background) target detection. Lower PRFs are usually used, which imply relatively high pulse compression ratios, as shown in Table 5.1. Scan rates are often slow with one bar taking 10 seconds.

A high range resolution profile can be used to recognize a ship just as with an aircraft.[72] It naturally has the same weakness previously mentioned, and the aspect or attitude must be known. If the attitude is known, then the major scatterers can be mapped into a range profile and correlated with the ship power return in each cell. An example of a ship range profile is shown in Figure 5.28. These profiles are usually generated in track when the profile is stabilized in range.

The wake of a surface ship or submarine near the surface provides a substantial cross section over time but requires surface stabilized integration over 10–100s of seconds.[16,45] Earth's surface stabilized integration can be done using a motion compensated doppler beam sharpening (DBS) mode.

Inverse SAR. A far more reliable method of ship recognition is inverse synthetic aperture radar (ISAR).[16,72] The basic notion is that the motion of a rigid object can be resolved into a translation and rotation with respect to the line of sight to the target. The rotation gives rise to a differential rate of phase change across the object. The phase history differences can be match filtered to resolve individual scatterers in a range cell. Conceptually, such a matched filter is no different than a filter used to match a phase-coded pulse compression waveform. This is the basis of all SAR, RCS range imaging, observed geometric target acceleration, turntable imaging, and ISAR.

A ship in open water exhibits roll, pitch, and yaw motions about its center of gravity (c.g.). For example, Figure 5.29 shows a rolling motion of ± 2.3° that might be exhibited by a ship in calm seas. The roll motion might have a period of 10 seconds. The motion of almost all the scatterers on a large combatant are moving in arcs of circles projected as segments of ellipses to a radar observer.[45] For a radar observer the change in range, dR, associated with a roll movement is a function of the height, *h*,

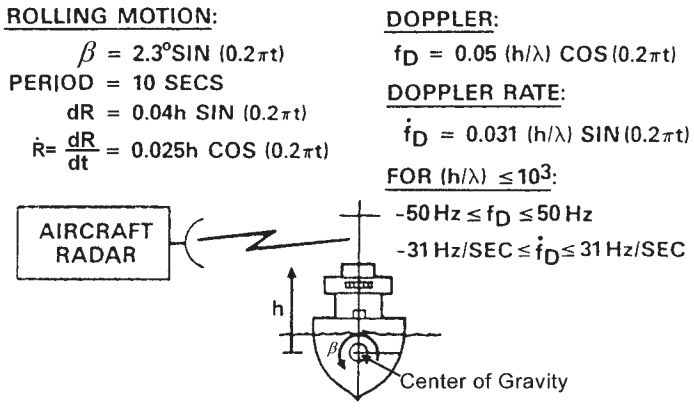


FIGURE 5.29 Inverse SAR notion[45]

of the scatterer above the center of gravity. The approximate range rate for each scatterer in rolling (pitch, yaw) motion at a height, h, is the time derivative of R shown in Figure 5.29. For a given desired cross range resolution with reasonable sidelobes, Δr_c, β must be equal to $\Delta r_c/\lambda$. For the example, 5 ft cross range resolution is obtainable with a 10-second observation time. The corresponding doppler and doppler rates are also given in Figure 5.29.

For a ship whose principal scatterers are less than 85 ft above the center of gravity, the dopplers will be in the range of ± 50 Hz at X band with a rate of change of up to ± 31 Hz/s. As long as the image resolution is not too great, each range-doppler bin can be match filtered using the hypothesized motion for each scatterer and an image can be formed on the ship. Each range bin may contain multiple scatterers from the ship in a given roll plane, and they may be distinguished by their differing phase history. However, scatterers in the pitch axis at the same range and roll height cannot be separated. Although pitch and yaw motions are slower, they also exist and allow separation in other similar planes.

Reasonably good images coupled with experienced radar operators allow recognition of most surface combatants. Recognition aids using prestored ship profiles allow identification to hull number in many cases. An example of a single ISAR image of a landing assault ship is given in Figure 5.30. The radar in this case is illuminating

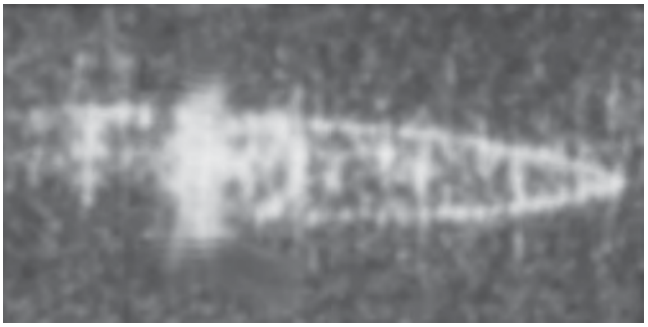


FIGURE 5.30 Single ISAR ship image[45]

the ship from the bow at 30 km and 6° grazing. The bright scatterers exhibit cross range sidelobes, which can be partially reduced by sensing large returns, then applying amplitude weighting and display compression, as has been done in this image. Integration of multiple ISAR images dramatically improves quality.

Air-to-Ground Ranging. Air-to-ground ranging is used most often for targeting of guns, dumb bombs, and missiles with short-range seekers against fixed or slow moving targets. The target is detected and designated in some other mode such as GMTI, DBS, SAR, or SSS. The designated target is tracked in range and angle to provide a more accurate distance and angle to the target. The tracking may be open or closed loop. The estimates are then provided to the weapon before and after launch. Depending on distance, another designator, such as a laser, and the radar may be alternately slaved to one another. Both the radar and the other designator may be subject to atmospheric refraction, especially at low altitudes, which is sometimes estimated and compensated.

Precision Velocity Update. Precision velocity update (PVU) is used for navigation correction to an inertial platform. Although GPS updates are commonly used to provide navigation in many situations, a military aircraft cannot depend solely on its availability. Furthermore, inertial sensors are used to fill in between GPS measurements even under the best circumstances. Inertial sensors are extremely good over short span times, but velocity drift is a major long-time error source, e.g., 1 km/h accumulates 16.6 m error per minute. A radar mode may require position to 0.1 km for proper operation.

PVU generally uses three or more antenna beam positions in which it makes a velocity measurement, as shown in Figure 5.31.[15] This mode directly emulates dedicated radar doppler navigators. There is a three-stage velocity measurement process. First, the surface is automatically acquired in range. Second, a fine range measurement is made, often using monopulse discriminants and range centroiding similar to that shown in Eq. 5.3. Third, a line-of-sight velocity measurement, V_{LOS}, using doppler and/or range rate, is made also using centroiding. Because terrain may be rising or falling at the illuminated patches giving rise to velocity errors, terrain slope is estimated and used to correct the estimated velocity.

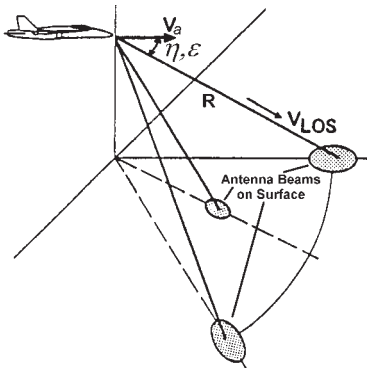


FIGURE 5.31 Precision velocity update concept[45]

A Kalman filter (a recursive filter that adaptively combines models of target measurements and of errors[10]) is employed to provide a better estimate of aircraft velocity. Although this procedure can be performed over land or water, sea currents make over-water measurements far less accurate. This velocity measurement provides in-flight transfer alignment of the various inertial platforms (aircraft, weapons, and radar). A set of outputs is provided to the mission management computer function, including North-East-Down (NED) velocity errors and estimates of statistical accuracies.

Sniff or Passive Listening. Most modes have a precursor subprogram called sniff, which looks for passive detections in a tentative operating channel before any radar emissions in that channel. The detections could be a friendly interferor, a jammer, or an inadvertent interferor such as a faulty civilian communications transponder.[99] This last example is the most common in the author's experience. It is not uncommon for a faulty transponder to appear as a million square meter target.

Doppler Beam Sharpening (DBS).[16,45,97] DBS is very similar to synthetic aperture radar (SAR) since both use the doppler spread across the antenna main beam to create higher resolution in the cross beam direction.[8,9,28,52] The principal difference is the amount of angular coverage, beam scanning, resolution, data gathering time, and accuracy of matched filtering in each range-doppler cell. A DBS map may take a second to gather over an angle of 70°. Depending on the angle from the aircraft velocity vector, a SAR map of a few feet resolution may take tens of seconds to gather at X band. DBS and SAR are compared in a qualitative way in Figure 5.32.

As the beam is positioned closer to the velocity vector, the doppler spread is smaller and so coherent dwell times must increase for the same resolution. Usually, there is a transition from shorter coherent processing intervals (CPIs) and longer post detection integrations (PDIs) to longer CPIs and shorter PDIs as the beam approaches the aircraft velocity vector. Near nose-on dwell times become prohibitive and the scan center is filled with real beam mapping. The real beam uses the same range resolution, but because returns from the entire beam are used, some amplitude equalization is required to provide uniform contrast and brightness across the whole map. Some effort is made

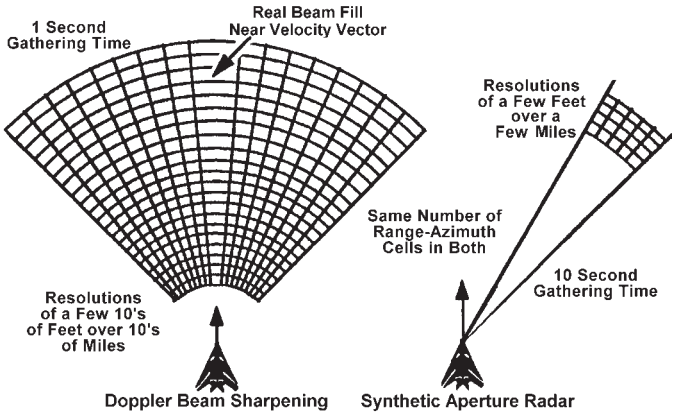


FIGURE 5.32 Doppler beam sharpening (DBS) comparison to SAR[9]

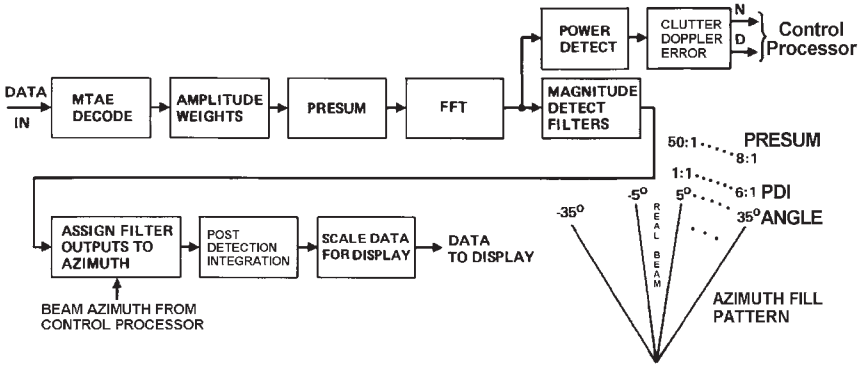


FIGURE 5.33 DBS processing (*adapted*[8]*; courtesy SciTech Publishing*)

to match filter (both in range closure and phase history) the doppler spread since iso-range and iso-dopplers are not close to orthogonal near the aircraft velocity vector (see Figure 5.8). SAR, on the other hand, is usually fully matched (relative to the desired resolution and phase history) in every range-doppler cell.

Figure 5.33 shows the signal processing that might be found in DBS mode. It consists of multiple time around echo (MTAE) suppression, amplitude weighting to improve sidelobes, presummation, an FFT filter bank, magnitude detection in each usable filter output, placement of each filter output in the correct ground stabilized location followed by post detection integration, and scaling for the display for constant brightness and dynamic range. Depending on grazing angle, ambiguous returns may compete with the region to be imaged. Often, a combination of sensitivity time control (STC) and pulse to pulse phase coding is used to reject multiple time around echoes (MTAE).[8,16,62,63] The amount of presummation (PRESUM) and post detection integration (PDI) as a function of beam position off the velocity vector is shown in the lower right of Figure 5.33. For each different angle, there is a different doppler spread across the beam. Therefore, in order to maintain a constant beam-sharpening ratio, different amounts of presumming must be used for each beam position. *Presumming* is the formation of an unfocussed synthetic beam (i.e., there is little or no attempt to match the exact phase history of surface points) inside the real antenna beam by what is essentially a lowpass filter. This would result in different target brightness and contrast if it were not compensated by applying a corresponding post detection integration (PDI) for each angle, as shown in Figure 5.33.

Multiple frequency looks are used to reduce speckle in the image and so several different frequencies are PDIed. The CPI is the presum ratio times the number of filter samples (128–800 is typical). Each CPI may have minor changes in the PRF to simplify processing and compensate for aircraft maneuvers. The aircraft may travel 1000 ft during the gathering time. There is considerable transport delay in most SAR and DBS processing; as a result, processed returns must be rectified (i.e., compensated for geometric distortion), motion compensated, and mapped into the proper space angle and range position. Since DBS usually maps a large area to provide overall ground situational awareness, the total range coverage is often covered in multiple elevation beams and range swaths. This is transparent to the operator but requires different PRFs, pulsewidths, filter shapes, and dwell times.

Although an MFAR contains a very stable time reference, uncertainties in the rate of change of terrain height, refraction, winds aloft, and very long coherent integration times force the measurement of the clutter doppler error versus predicted frequency to maintain proper focus and bin registration, as shown in the upper right in Figure 5.33. A similar function is performed in SAR as well.

Synthetic Aperture Radar. As is the case for DBS, SAR is a multirate-filtering problem, i.e., a cascade of filters in which the input sampling rate is higher than the output sampling rate, as shown in Figure 5.35, which requires very careful attention to range and azimuth filter sidelobes. Typically, the spacing of individual pulses on the ground is chosen to be much closer than the desired ultimate resolution. This allows linear range closure and phase correction since each point on the surface moves a significant fraction of a range cell pulse to pulse.[38,40–42,66,69,91] The input signal, point A in Figure 5.34, is shown as a spectrum at A, folded about the PRF on the left in Figure 5.35.

Subsequently, presummation is applied, which forms an unfocussed synthetic beam or filter inside the main-beam ground return (point B in Figure 5.34), which improves azimuth sidelobes and narrows the spectrum, as suggested in the center graph shown in Figure 5.35. The presummer output is resampled at a lower rate, f_S, consistent with acceptable filter aliasing. Then, range pulse compression is performed, assuming the transmitted pulse is very long compared to the range swath. If chirp (linear FM) is used, part of the "stretch" pulse compression processing is performed in the range compression function with the remainder performed in polar format processing. The dechirped and partially filtered or compressed output, shown at point C in Figure 5.34, may be resampled again at a new f_S, as indicated in the right graph shown in Figure 5.35, point C. In any case, azimuth variable phase adjustment and bin mapping (which compensates for changes in measurement space angles and range closure since significant motion occurs during the data gathering time) must be performed before azimuth filtering (sometimes called compression because it is similar to phase matched pulse compression). The output of azimuth compression is shown at point C. The complex SAR output map must be checked for depth of focus and usually requires autofocus since both atmospheric effects and locally rising or falling terrain may cause defocusing. Subsequent to refocusing, the map is magnitude detected and histogram averaged to maintain uniform brightness. The map is integrated with other looks, which requires

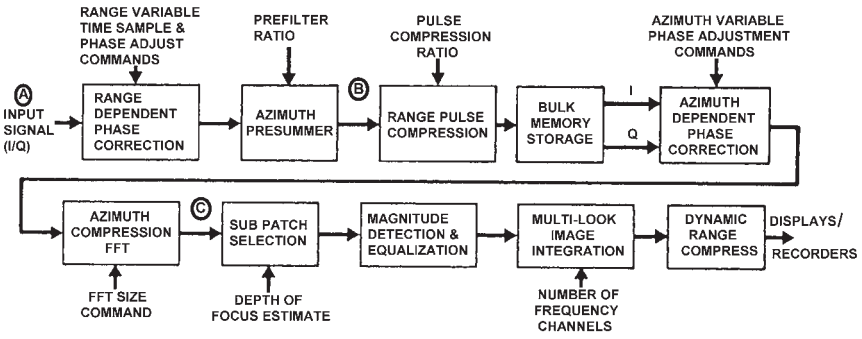


FIGURE 5.34 SAR processing (*adapted[8]; courtesy SciTech Publishing*)

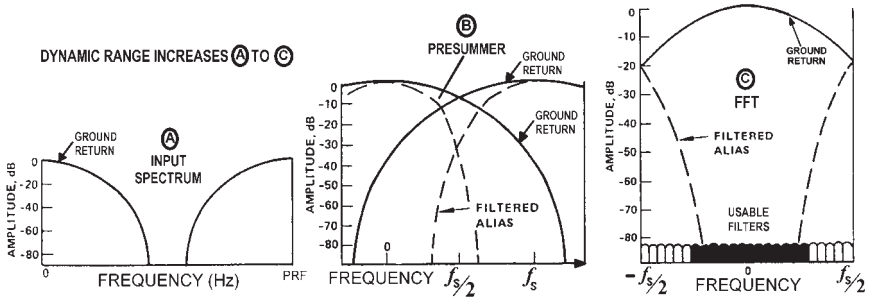


FIGURE 5.35 SAR Multirate Filtering (*adapted*[8]*; courtesy SciTech Publishing*)

geometrical correction and motion compensation. The total map dynamic range can easily be greater than 60 dB. The typical cockpit display is limited to 15–25 dB and dynamic-range compression, such as converting map amplitudes into their logarithms, is often performed.

DBS or SAR PRF, Pulse Length and Compression Selection. For each SAR or DBS geometry, the transmitted pulse width, pulse repetition interval, and pulse compression ratio must be calculated. One possible set of selection criteria is given in Eq. 5.4.[45]

Usually, the last range ambiguity before the range swath is chosen to be outside the main beam, far enough to be at least 20 dB down, including R^4 effects. Often in SAR, the transmitted pulse is much larger than the range swath, R_{swath}. Clearly, in each of the cases, the nearest integral clock interval and nearest convenient pulse compression ratio is selected because the values in Eq. 5.4 will be clock integers only by coincidence.

Pulse Repetition Interval (*PRI*):

$$\frac{\lambda}{2 \times V_a \times U_0 \times B_{\text{az}} \times \sin(\theta)} \geq PRI \geq \frac{2 \times (R_1 - R_{\min} + R_{\text{swath}} + R_p)}{c}$$

Pulse Width: $R_p \leq \text{Duty}_{\max} \times PRI \times c$
Minimum Allowable Ambiguous Range:
$R_{\min} \approx h \times \csc(\varepsilon + U_1 \times B_{\text{el}}/2)$
Range Swath is Geometry and Instrumentation Dependent:
$R_{\text{swath}} \leq h \times [\csc(\varepsilon - B_{\text{el}}/2) - \csc(\varepsilon + B_{\text{el}}/2)]$
and $R_{\text{swath}} \leq R_{\text{maxswath}}$ (5.4)

where: λ is transmitted wavelength, h is the aircraft altitude,
B_{az} & B_{el} are the azimuth & elevation half power beamwidths,
θ & ε are the angles between the velocity vector and antenna beam center,
R_1 is the distance to the first range bin, V_a is the aircraft velocity,
R_{swath} is the range swath length, R_{maxswath} is maximum instrumented range swath,
$R_{\min}$ is the range to the closest allowable ambiguity, $\text{Duty}_{\max}$ is allowable duty ratio,
R_p is the transmitted pulse length in distance units, c is the velocity of light,
U_0, U_1 are beamwidth multipliers at predefined power rolloff.

For example, assume $V_a = 300$ m/s, $\lambda = 0.03$ m, $h = 5000$ m, $\theta = 0.5$, $\varepsilon = 0.1$, B_{az}, $B_{el} = 0.05$, U_0, $U_1 = 2.3$, $R_{swath} = 2$ km, $R_{min} = 32$ km, desired mapping range, $R_1 = 50$ km, $\text{Duty}_{max} = 0.25$, selecting a first guess for $R_p = 8000$ m; then $186 < PRI < 906$ μsec, R_{min} is the equivalent of 213 μsec, and the next allowable ambiguity would be past the swath at 400 μsec; therefore, a PRI of 213 or 400 μsec could be used with a transmitted pulse of approximately 50 or 100 μsec respectively.

Ground Moving Target Indication (GMTI) and Track (GMTT). GMTI is the detection and acquisition of ground moving targets. GMTI and GMTT radar modes have a different set of challenges. First, target detection is usually the easy part; the RCS of most anthropogenic objects and many natural moving targets is large (10–1000 m^2). Unfortunately, there are many stationary objects with moving parts such as ventilators, fans, water courses, and power lines that lead to apparent false alarms.[93] Often slow-moving vehicles have fast-moving parts (e.g., helicopters and agricultural irrigators).

Most areas have large numbers of vehicles and scatterers that could be vehicles. It is typical to have up to 20,000 bona fide GMTs in the field of view. Processing capacity must be adequate to handle and discriminate thousands of high *SNR* threshold crossings and hundreds of moving targets of interest. Usually multi-hypothesis tracking[10] filters will be following several hundred GMTs of interest simultaneously. In most cases, all targets must be tracked and then recognized on the basis of doppler spectrum (helicopters vs. wheeled vehicles vs. tracked vehicles vs. scanning antennas), rate of measured location change (ventilator locations don't change), and consistent trajectory (e.g., 60 mph where there are no roads is improbable for a surface vehicle).[94] In addition, vehicles of interest may have relatively low radial velocities requiring endoclutter processing[16] (i.e., far enough inside main-beam clutter that detection is limited for doppler only filtering).

A processing block diagram for GMTI is shown in Figure 5.36. Although there are alternate ways to perform endoclutter processing, a multiple phase center–based

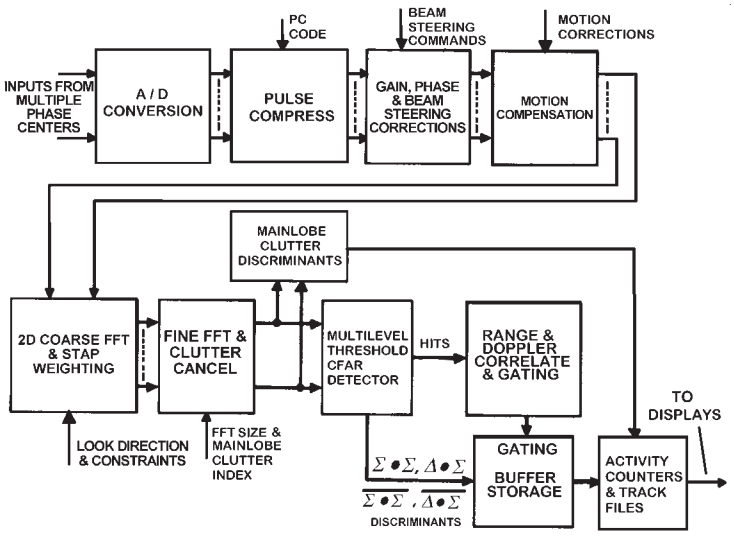


FIGURE 5.36 Ground moving target detection processing (*adapted[8]; courtesy SciTech Publishing*)

processing scheme is given in Figure 5.36. Multiple channels or phase centers are digitized and pulse compressed. Periodic calibration signals are used to create a gain, phase and beam steering correction table for all frequencies, antenna beam steering, and channels, which are then applied to the digitized measurements in each channel. Motion compensation to a fraction of a wavelength for platform maneuvers or deviations is applied to the data. A coarse two-dimensional FFT is performed followed by space-time adaptive calculations, and filter weighting is applied to reject some clutter and jamming. High resolution doppler filtering is performed in a conventional FFT,[90,96,97] perhaps with DPCA clutter cancellation. Doppler filter outputs are used to form main-beam clutter error discriminants for precisely measuring doppler center frequency to provide fraction of wavelength motion compensation. Main-beam clutter is not in the same frequency location for each range bin, and so filter output order must be adjusted to present a common input to the threshold detector. The doppler filter bank outputs also are applied to a multilevel threshold detector for ground moving target detection similar to those described in "Ground Moving Target Thresholding." Sum and difference discriminant functions are formed and stored in buffer storage for each detected moving target to improve target tracking and geolocation accuracy.

Often PRFs are ambiguous in both range and doppler but unambiguous inside the main beam and near sidelobes (i.e., there is only one range or doppler ambiguity interval in the main beam and near sidelobes). PRF selection is similar to A-A MPRF. Usually fewer PRFs are used; four or five are typical.[97] A range ambiguity may be in the main beam at low grazing angles. Two out of four or three out of five is usually the final detection criteria. PRFs typically are 4–8 kHz. Coded waveforms are often used to reject ambiguous returns outside the antenna main beam that compete with the region of interest. A 10 ft range cell size is often used to match the smallest vehicle of interest and to reduce background clutter. Ground moving target recognition may require 0.25 ft resolution. Antenna illumination must be ground stabilized since the aircraft will engage in both intentional and unintentional maneuvers.[94]

Ground Moving Target Thresholding. The typical multilevel threshold has several unique features. In addition to the obvious alert-confirm properties (a double thresholding method in which a lower first threshold nominates radar returns as possible targets to be confirmed by a return observation with a higher threshold), it also uses multiple phase center discriminants as well as near sidelobe threshold multipliers.[94] Even with STAP, the non-gaussian nature of clutter requires higher thresholds in the main beam and near sidelobes.[98] Threshold crossings are correlated in range and doppler and buffered along with corresponding phase center discriminants, which are presented to tracking filters or activity counters.

There are three regions of thresholding: main-beam clutter-limited detection, near sidelobe clutter-limited detection, and thermal-noise-limited detection. Near surface targets of interest will often have radial velocities of a few miles per hour for long periods of time, which forces the detection of ground moving targets well into main-beam clutter. Phase monopulse, DPCA, or STAP processing allows the first order cancellation of clutter for many slow-moving targets. Unfortunately, clutter does not always have well-behaved statistical tails, and to maintain a constant false alarm rate, the threshold must be raised for endoclutter targets. The output of the doppler filter bank might be thought of as a two dimensional range-doppler image. There will still be parts of main-beam clutter that are completely discarded except for motion compensation because clutter cancellation is inadequate.[98]

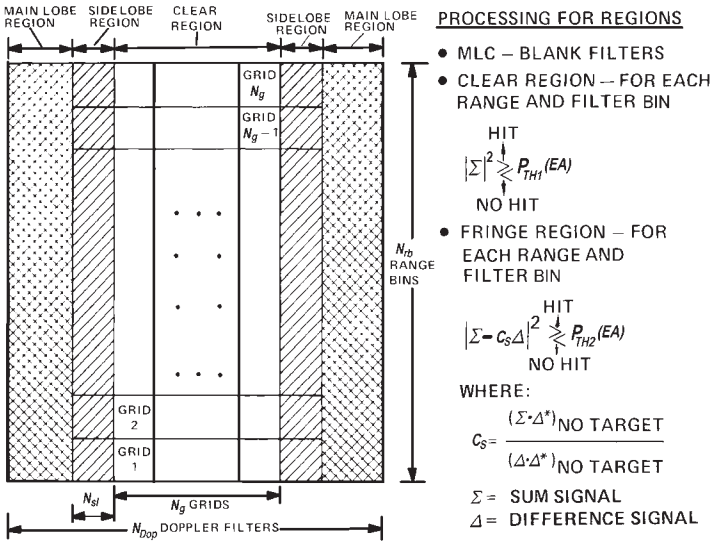


FIGURE 5.37 Multiregion GMT thresholding[8] (*Courtesy SciTech Publishing*)

An example thresholding scheme based on these concepts is shown in Figure 5.37. The range-doppler space is broken up into a grid of range bins and doppler filters, as shown in the figure. Each cell in the grid might be 64 × 64 range-doppler bins with 256 grid cells total. Some grid locations close to main-beam clutter (MLC in figure) are used for forming main-beam clutter discriminants only and are otherwise discarded. The bins (64^2 in the example) in each grid cell are ensemble averaged (EA) in sum and difference channels. The power in each bin in a grid cell in the clear (thermal noise limited) region is compared to a threshold, $P_{TH1}(EA)$, which is a function of the EA in that grid cell. In the endoclutter near sidelobe region, a discriminant, C_s, is formed and used to provide additional clutter cancellation prior to thresholding. Again, the threshold, $P_{TH2}(EA)$, is a function of the EA in that grid cell and a priori knowledge of the clutter statistics. Although only one threshold is described, two are actually used before hits and their corresponding discriminants are passed to the track files. All low threshold hits are passed to activity counters. As complex as this thresholding scheme seems to be, it is very detection power efficient.

Typical GMT Weapon Delivery. As mentioned previously, missile guidance requires tracking of both targets and missiles (also bullets in gun laying radar; *gun laying* is a term invented by the U.K during WWII for radar pointing of antiaircraft guns). Range accuracy is at least an order of magnitude better than angle accuracy. Some method must be used to improve angle accuracy for weapon delivery. An example processing diagram for GMT weapon delivery is shown in Figure 5.38. In this case, three different classes of target or missile are tracked. A single waveform may be used to track stationary, endo- and exoclutter moving targets, and missiles or bullets. Each class of return, based on its range and doppler location, is separately tracked and geolocated.[94]

There are several common types of geolocation; many of them are based on using either DTED or cartographic data. One method using cartographic data is shown

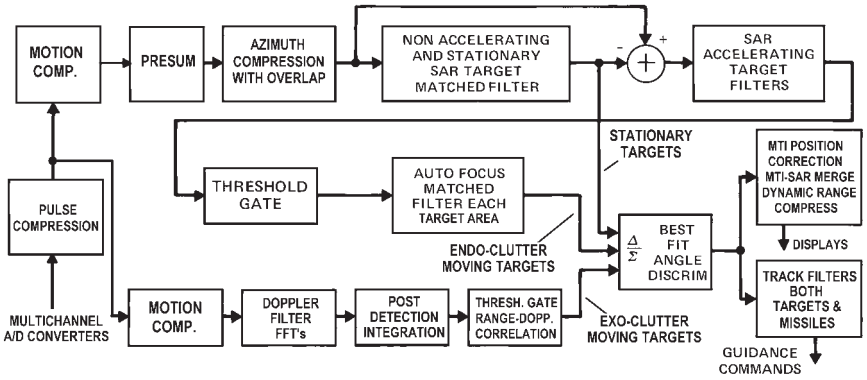


FIGURE 5.38 Typical GMT weapon guidance (*adapted*[8]*; courtesy SciTech Publishing*)

in Figure 5.39. An error ellipse and its corresponding eccentricity are calculated for each target. If the eccentricity is less than some arbitrary threshold (e.g., 0.866 relative), the minimum perpendicular distance is calculated for road segments inside the 3-sigma ellipse. As shown in the figure, the perpendicular intercept may not lie inside the road segment and will be discarded. The minimum distance for valid road segment distances will be selected as the GMT location. If the eccentricity is greater than the threshold, the road segments that have a major ellipse axis intercept inside 3 sigma are compared, and the minimum distance is selected. Obviously, some other screening must also be applied. For example, some roads cannot support high speeds and tanks do not have to be on roads.

A common SAR-MTI display may be presented to the operator. In addition, guidance commands or errors are derived from the measurements and provided to downlinks to either missiles on the fly or gun directing computers for the next rounds. Short-term coherent change detection may be used to separate stationary targets from slow-moving endoclutter targets. Short-term coherent change detection is a method in which two coherent SAR maps taken within a few hours of one another at the same frequency are registered and cross-correlated pixel by pixel. The fast-moving target category usually includes both targets and bullets or missiles.

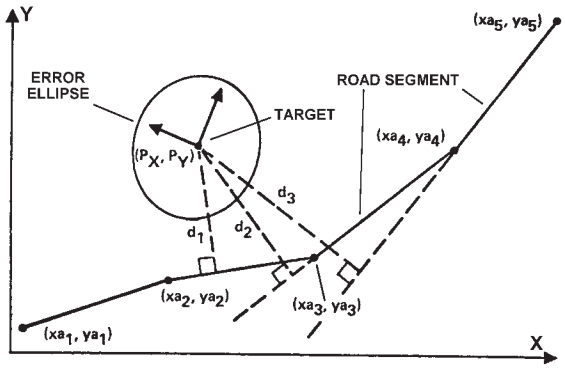


FIGURE 5.39 Cartographic-assisted GMT geolocation[94]

Missile Performance Assessment, Track, and Update. Missile midcourse guidance usually consists of assessing the missile performance, measurement of the target and missile location, prediction of the path of each, and updating the resulting data to the missile for the best future intercept of the target. It may also include the most current estimate of the target type and attitude for best fuzing. The missile usually sends data about its state of health, ownship measurements, remaining fuel amounts, and target acquisition, if any.[74] When the missile is close to the data link aircraft (which may or may not be the launching platform), communication is often through an aperture other than in the main MFAR. As the distance gets greater, the primary MFAR aperture is used. As the data link aircraft maneuvers, the aperture that has the largest projected area in the direction of the missile is used. The bandwidth to the missile is very low and can be redundant and highly encrypted to provide good antijam (A/J) protection. If it contains imagery, the uplink bandwidth from the missile is relatively large and will have comparatively lower A/J performance. An adaptive MFAR primary aperture can improve a wider band missile uplink A/J if the jammer is offset from the target. At the missile end, the missile antenna can have jammer nulling to improve downlink A/J.[43,85]

AGC, Calibrate and Self-Test. Usually at the beginning of a new mode, the end of each scan bar, or once per second, the calibrate and self-test subprogram is invoked by the operational flight program (OFP) executive. A sequence of subroutines is executed that measures phase and gain unbalance between channels using a signal injected on the antenna. This is usually done over a range of input amplitudes, frequencies, and AGC settings because of the nonlinear characteristics of most RF front ends. Also for modes like TF/TA, a full set of off-angle diagnostics is performed, which tests the integrity of the entire measurement, processing, and flight control chain often enough to keep the probability of a failure-induced crash per flight below 10^{-6} in the presence of jamming or component failures.

In addition, there are initiated built-in tests at two levels: an operational readiness test performed as part of mission initiation and a fault isolation test performed by the maintenance crew in response to an operator deficiency report. Both tests take longer and are more exhaustive. In the best case, the specific flight line or a first-level maintenance replaceable assembly is identified with high probability. Such assemblies are then sent to a depot for replacement, repair, failure tracking, and/or reclamation. For assemblies that have a very low failure rate, it is usually cheaper to replace and reclaim rather than repair even when the assembly is very expensive.

REFERENCES

Short course notes and other papers can usually be obtained from the authors or the course sponsor for a small fee. All of the authors' papers referenced are available in Adobe Acrobat format subject only to copyright restrictions by e-mail request: davidlynchjr@ieee.org and carlo.kopp@iinet.net.au.

1. C. Kopp, "Active electronically steered arrays," 2002, http://www.ausairpower.net.
2. Joint Advanced Strike Technology Program, "Avionics architecture definition 1.0," U.S. DoD public release, unlimited distribution and use, pp. 9, 11, 31,32, 1994.
3. D. Eliot (ed.), *Handbook of Digital Signal Processing*, San Diego, CA: Academic Press, 1987, pp. 364–464, 527–589, 590–593, 594–631.
4. L. Tower and D. Lynch, "Pipeline High Speed Signal Processor," U.S. Patent 4,025,771, 5/24/1977.

5. L. Tower and D. Lynch, "System for Addressing and Address Incrementing of Arithmetic Unit Sequence Control System," U.S. Patent 4,075,688, 2/21/1978.
6. L. Tower and D. Lynch, "Pipelined microprogrammable control of a real time signal processor," in *IEEE Micro6 Conference*, June 1973, p. 175.
7. D. Lynch, "Radar systems for strike/fighter aircraft," presented at *AOC Third Radar/EW Conference Proceedings*, Unclassified paper in classified proceedings available from author by request, February 12–13, 1997.
8. D. Lynch, *Introduction to RF Stealth,* Raleigh, NC: SciTech Publishing, 2004, pp. 446–451, 82–84, 198–221, 467–470 492–501, 504–531.
9. D. Lynch et al., "Advanced avionics technology," Evolving Technology Institute Short Course Notes, November 1994.
10. S. S. Blackmun, *Multiple Target Tracking with Radar Applications*, Dedham, MA: Artech House, 1986, pp. 25–44, 281–298.
11. D. A. Fulghum and D. Barrie, "Radar becomes a weapon," *Aviation Week & Space Technology*, pp. 50–52, September 5, 2005.
12. Image Courtesy Raytheon Company, cleared for public release, 265-SPR127.05.
13. M. Streetly, *Radar and Electronic Warfare Systems*, 1999–2000, 11th Ed., Coulsdon, Surrey, UK: Janes Information Group, 1999, pp. 250–254.
14. R. Nitzberg, *Radar Signal Processing and Adaptive Systems*, Norwood, MA: Artech House, 1999, pp. 199–202, 207–236, 267–290.
15. W. K. Saunders, "CW and FW radar"; F. M. Staudaher, "Airborne MTI"; W. H. Long, D.H. Mooney, and W. A. Skillman, "Pulse doppler radar"; R. J. Serafin, "Meteorological radar," *Radar Handbook*, 2nd Ed., M. Skolnik (ed.), New York: McGraw Hill, 1990, pp. 14.37–14.39, 16.8–16.28, 17.33–17.35, 23.5–23.13.
16. P. Lacomme, J-P. Hardange, J-C. Marchais, and E. Normant, *Air and Spaceborne Radar Systems: An Introduction*, Norwich, NY: William Andrew Publishing, 2001, pp. 329–335, 371–425, 171–176, 469.
17. J. Davis, "Sun intros eight–core processor," *Electronic News*, Reed Elsevier, November 14, 2005.
18. Altera Corporation, "Stratix II FPGA's," November 2005, http://www.altera.com.
19. D. A. Fulghum, "Deep look," *Aviation Week and Space Technology*, January 17, 2005.
20. D. A. Fulghum, "Future radar," *Aviation Week and Space Technology*, October 4, 2004.
21. M. Peck and G. W. Goodman, Jr., "Agile radar beams," *C4ISR Journal*, pp. 22–28, May 2005.
22. "Raytheon's APG-79 AESA radar for the F/A-18 Super Hornet sets a new standard as it delivers multiple JDAMs simultaneously on target," *MarketWatch*, December 5, 2005.
23. M. Selinger, "U.S. Navy eyes 'growth plan' for Super Hornet's AESA radar," *Aerospace Daily and Defense Report*, December 6, 2005.
24. R. E. Hudson, S. O. AKS, P. P. Bogdanovic, and D. D. Lynch, "Method and System for Reducing Phase Error in a Phased Array Radar Beam Steering Controller, U.S. Patent 4,924,232, 5/8/1990.
25. R. Hill, D. Kramer, and R. Mankino, "Target Detection System in a Radar System Employing Main and Guard Channel Antennas," U.S. Patent 3875569, 4/1/1975.
26. R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*, New York: John Wiley & Sons, 1980, pp. 78–279.
27. R. Klemm, "Adaptive airborne MTI: An auxiliary channel approach," *IEE Proceedings*, vol. 134, part F, no. 3, p. 269, 1987.
28. S. Aks, D. D. Lynch, J. O. Pearson, and T. Kennedy, "Advanced modern radar," Evolving Technology Institute Short Course Notes, November 1994.
29. Work performed by L. Griffiths and C. Tseng, "Adaptive array radar project review," Hughes Aircraft IR&D, performed at USC, July 18, 1990.
30. C. Ko, "A fast adaptive null-steering algorithm based on output power measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 3, pp. 717–725, July 1993.

31. H. Wang, H. Park, and M. Wicks, "Recent results in space-time processing," in IEEE National Radar Conference 1994, pp. 104–109.
32. J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Laboratory Report 1015, approved for unlimited public distribution.
33. N. M. Greenblatt, J. V. Virts, and M. F. Phillips, "F-15 ESA medium PRF design," Hughes Aircraft IDC No. 2312.20/804, January 9, 1987, unclassified report.
34. D. Schleher, "Low probability of intercept radar" in *IEEE International Radar Conference*, 1985, p. 346.
35. E. Carlson, "Low probability of intercept techniques and implementations," in *IEEE National Radar Conference*, 1985, p. 51.
36. Groger, "OLPI-LPI radar design with high ARM resistance," in *DGON 7th Radar Conference* 1989, p. 627.
37. D. Lynch, "Real time radar data processing," presented at IEEE Solid State Circuits 4.10 Committee, Digital Filtering Meeting, New York, October 30, 1968.
38. D. Craig and M. Hershberger, "FLAMR operator target/OAP recognition study," Hughes Aircraft Report No. P74-524, January 1975, Declassified 12/31/1987.
39. D. C. Schleher, *Electronic Warfare in the Information Age*, Norwood, MA: Artech House, 1999, pp. 279–288, 133–199.
40. S. Hovanessian, *Introduction to Synthetic Array and Imaging Radars*, Dedham, MA: Artech House, 1980, Chapter 5.
41. J. Curlander and R. McDonough, *Synthetic Aperture Radar Systems and Signal Processing*, New York: Wiley & Sons, 1991, pp. 99–124, 427–535.
42. J. Kovaly, *Synthetic Aperture Radar*, Dedham, MA: Artech House, 1976, pp. 72–79, 118–123, 249–271.
43. B. Lewis, F. R. Kretschmer, and W. W. Shelton, *Aspects of Radar Signal Processing*, Dedham, MA: Artech House, 1986, pp. 267–290.
44. M. Radant, D. Lewis, and S. Iglehart, "Radar sensors," UCLA Short Course Notes, July 1973.
45. D. Lynch, J. O. Pearson, and E. Shamash, "Principles of Modern radar," Evolving Technology Institute Short Course Notes, June 1988.
46. J. Frichel and F. Corey, "AN/APG-67 Multimode Radar Program," in *IEEE NAECON 1984*, p. 276.
47. R. Nevin, "AN/APG-67 multimode radar performance evaluation," in *IEEE NAECON 1987*, p. 317.
48. D. Lynch, "SLOSH filter processing," presented at IEEE AU Symposium on Digital Filters, Harriman, NY, January 1970.
49. Treffeisen et al., "Obstacle Clearance System for Aircraft," U.S. Patent 3,530,465, 9/22/1970.
50. *International Defense Review- Air Defense Systems*, Geneva, Switzerland: Interavia, 1976, pp. 61–103.
51. C. Kopp, "Missiles in the Asia-Pacific," *Defence Today*, http://www.ausairpower.net/DT-Missile-Survey-May-05.pdf.
52. G. Stimson, *Introduction to Airborne Radar*, 2nd Ed., Mendham, NJ: SciTech Publishing, 1998, pp. 355–381, 463–465, 431–437.
53. J. Clarke, "Airborne radar" Parts 1 & 2, *Microwave Journal*, p. 32 and p. 44, January 1986 and February 1986.
54. E. Aronoff and D. Kramer, "Recent developments in airborne MTI radars," Hughes Aircraft Report, presented at IEEE Wescon 1978.
55. D. Kramer and G. Lavas, "Radar System with Target Illumination by Different Waveforms," U.S. Patent 3866219, 2/11/1975.
56. D. Mooney, "Post-Detection STC in a Medium PRF Pulse Doppler Radar," U.S. Patent 4095222, 6/13/1978.
57. E. Frost and L. Lawrence, "Medium PRF Pulse Doppler Radar Processor for Dense Target Environments," U.S. Patent 4584579, 4/22/1986.
58. W. Long and K. Harriger, "Medium PRF for the AN/APG-66 radar," in *IEEE Proceedings*, vol. 73, no. 2, p. 301.

59. E. Aronoff and N. Greenblatt, "Medium PRF radar design and performance," Hughes Aircraft Report, presented at IEEE National Radar Conference 1975.
60. H. Erhardt, "MPRF processing functions-issue 2," Hughes Aircraft IDC, October, 18, 1977, unclassified report.
61. J. Kirk, "Target Detection System in a Medium PRF Pulse Doppler Search/Track Radar Receiver," U.S. Patent 4079376, 3/14/1978.
62. K. Gerlach, "Second time around radar return suppression using PRI modulation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-25, no. 6, pp. 854–860, November 1989.
63. L. Durfee and W. Dull, "MPRF Interpulse Phase Modulation for Maximizing Doppler Clear Space," U.S. Patent 6518917, 2/11/2003.
64. S. Hovanessian, "An algorithm for calculation of range in multiple PRF radar," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-12, no. 2, March 1976, pp. 287–290.
65. G. Morris, *Airborne Pulse Doppler Radar*, Norwood, MA: Artech House, 1988.
66. R. Schlolter, "Digital realtime SAR processor for C & X band applications," in *IGARSS 1986*, Zurich, vol. 3, p. 1419.
67. R. Klemm, "Airborne MTI via digital filtering," in *IEE Proceedings*, vol. 136, part F, no. 1, 1989, p. 22.
68. Technology Service Corp., "Adaptar space-time processing in airborne radars," TSC-PD-061-2, February 24, 1971, unclassified report.
69. D. Lynch, "Signal processor for synthetic aperture radar," presented at SPIE Technical Symposium East 1979, paper no. 180-35.
70. J. Harmon, "Track before detect performance for a high PRF search mode," in *IEEE National Radar Conference* 1991, pp. 11–15.
71. J. R. Guerci, *Space-Time Adaptive Processing for Radar*, Norwood, MA: Artech House, 2003, pp. 6, 11–28, 51–74.3.
72. P. Tait, *Introduction to Radar Target Recognition*, Bodmin, Cornwall, UK: IEE, 2005, pp. 105–217, 317–347.
73. P. Peebles, *Radar Principles*, New York: John Wiley & Sons, 1998, pp. 318–349, 599–614.
74. E. Eichblatt, *Test and Evaluation of the Tactical Missile*, Washington, DC: AIAA, 1989, pp. 13–39, 52–54.
75. R. Macfadzean, *Surface Based Air Defense System Analysis*, self-published, 1992 & 2000, pp. 213–243.
76. M. Robin and M. Poulin, *Digital Television Fundamentals*, 2nd Ed., New York: McGraw-Hill, 2000, pp. 345–425.
77. W. Pratt, *Digital Image Processing*, New York: Wiley & Sons, 1978, pp. 662–707.
78. M. Simon, J. Omura, R. Scholtz, and B. Levitt, "Low probability of intercept communications," Chapter 4 in *Spread Spectrum Communications Handbook*, New York: McGraw-Hill, 1994, pp. 1031–1093.
79. W. Gabriel, "Nonlinear spectral analysis and adaptive array superresolution techniques," NRL Report 8345, 1979, approved for unlimited public distribution.
80. J. Asenstorfer, T. Cox, and D. Wilksch, "Tactical data link systems and the Australian defense force (ADF)-technology developments and interoperability issues," Defense Science and Technology Organisation Report, DSTO-TR-1470, approved for public release.
81. C. Kopp; "The properties of high capacity microwave airborne ad hoc networks," Ph.D. dissertation, Monash University, Melbourne, Australia, October 2000.
82. J. Katzman, *Defence Industry Daily*, http://www.defenseindustrydaily.com/2005/12/elec-tricks-turning-aesa-radars-into-broadband-comlinks/index.php.
83. C. Nakos, S. Baker, J. J. Douglass, and A. R. Sarti, "High speed data link," Australia Patent PCT/AU97/00255, WO 97/41450, DSTO Tactical Surveillance Systems Division, Salisbury, Australia, November 1997.
84. D. A. Fulghum, "See it, jam it, kill it" *Aviation Week & Space Technology*, pp. 24, 25, May 30, 2005.

85. D. C. Schleher, *Introduction to Electronic Warfare*, Dedham, MA: Artech House, 1986, pp. 280–283, 109–128.
86. Case, Jr. et al., "Radar for Automatic Terrain Avoidance," U.S. Patent 3,815,132, 6/4/1974.
87. H. L. Waruszewski, Jr., "Apparatus and Method for an Aircraft Navigation System Having Improved Mission Management and Survivability Capabilities," U.S. Patent 5,086,396, 2/4/1992.
88. Barney et al., "Apparatus and Method for Adjusting Set Clearance Altitude in a Terrain Following Radar," U.S. Patent 4,760,396, 7/26/1988.
89. R. Jaworowski, "Outlook/specifications military aircraft," *Aviation Week and Space Technology*, pp. 42–43, January 17, 2005.
90. F. Harris and D. Lynch, "Digital signal processing and digital filtering with applications," Evolving Technology Institute Short Course Notes, 1971–1983, pp. 366, 744–748, February 1978.
91. R. Fabrizio, "A high speed digital processor for realtime SAR imaging," in *IGARSS 1987*, Ann Arbor MI, vol. 2, p. 1323.
92. T. Cullen and C. Foss (eds.), *Janes Land-Based Air Defence 2001–2002*, Coulsdon, Surrey, UK: Janes Information Group, 2001, pp. 129–134.
93. R. Klemm, "New airborne MTI techniques," in *International Radar Conference London*, 1987, p. 380.
94. "Pave mover TAWDS design requirements," Hughes Aircraft Specification, November 1979, unclassified, unlimited distribution.
95. J. Pearson, "FLAMR signal to noise experiments," Hughes Aircraft Report No. P74-501, December 1974, declassified 12/31/1987.
96. E. O. Brigham, *The Fast Fourier Transform*, New York: Prentice Hall, 1974, pp. 172–197
97. D. Lynch, et al., "LPIR phase 1 review," Hughes Aircraft Report, 1977, unclassified report.
98. J. O. Pearson, "Moving target experiment and analysis," Hughes Aircraft Report No. P76-432, pp. 5–15, 22–35, December 1976, declassified 2/28/94.
99. K. Rogers, "Engineers unlock mystery of car-door device failures," *Las Vegas Review Journal*, August 19, 2004, p. 1B.

Chapter 6
Radar Receivers*

Michael E. Yeomans

Raytheon Company

6.1 THE CONFIGURATION OF A RADAR RECEIVER

The function of a radar receiver is to amplify, filter, downconvert, and digitize the echoes of the radar transmission in a manner that will provide the maximum discrimination between desired echo signals and undesired interference. The interference comprises not only the self noise generated in the radar receiver but also the energy received from galactic sources, neighboring radars and communication equipment, and possibly jammers. The portion of the radar's own radiated energy that is scattered by undesired targets (such as rain, snow, birds, insects, atmospheric perturbations, and chaff) may also be classified as interference and is commonly categorized as clutter. Where airborne radars are used for altimeters or mapping, other aircraft are undesired targets, and the ground is the desired target. In the case of weather radars, ground, buildings, and aircraft are clutter, and rain or snow is the desired target. More commonly, radars are intended for detection of aircraft, missiles, ships, surface vehicles, or personnel, and the reflection from weather, sea, or ground is classified as clutter interference.

Although the boundaries of the radar receiver are somewhat arbitrary, this chapter will consider those elements identified in Figure 6.1 as the receiver. The radar exciter generates the transmit waveforms as well as local oscillator (LO), clock, and timing signals. Since this function is usually tightly coupled to a radar receiver, it is also shown in Figure 6.1 and will be discussed in this chapter. The purpose of Figure 6.1 is to illustrate the functions typical of a modern radar receiver and exciter.

Virtually all radar receivers operate on the superheterodyne principle shown in Figure 6.1. Through this architecture, the receiver filters the signal to separate desired target signals from unwanted interference. After modest RF amplification, the signal is shifted to an intermediate frequency (IF) by mixing with a local-oscillator (LO) frequency. More than one conversion stage may be necessary to reach the final IF without encountering serious image- or spurious-frequency problems in the mixing process. The superheterodyne receiver varies the LO frequency to follow any desired tuning variation of the transmitter without disturbing the filtering at IF. This simplifies the filtering operation as the signals occupy a wider percentage

* This chapter incorporates material written by John W. Taylor, Jr. for the first and second editions and updated by Michael Yeomans for this edition.

bandwidth at the IF frequency. These advantages have proven to be so significant that competitive forms of receivers have virtually disappeared.

In conventional antenna systems, the receiver input signal is derived from the duplexer, which permits a single antenna to be shared between transmitter and receiver. In active array systems, the receiver input is derived from the receive beam-forming network. Active array antennas include low-noise amplifiers prior to forming the receive beams; although these are generally considered to be antenna rather than receiver components, they will be discussed in this chapter.

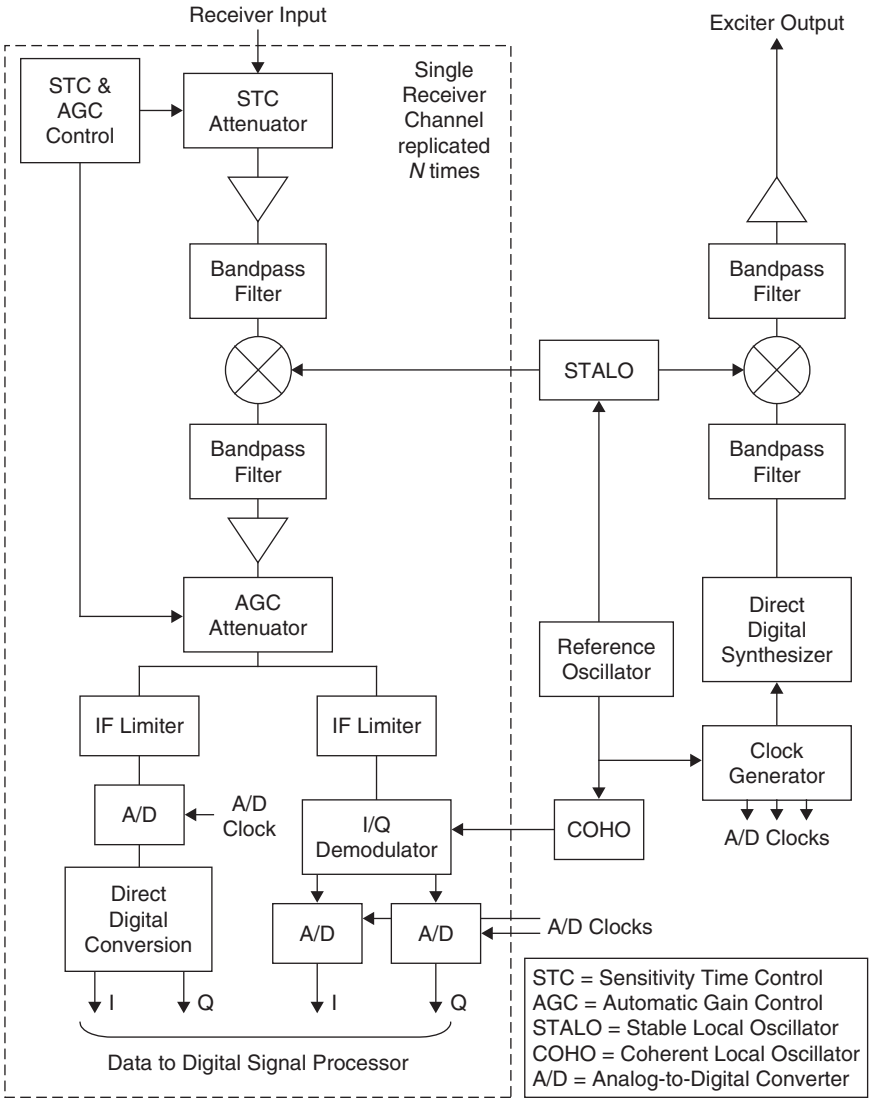


FIGURE 6.1 General configuration of a radar receiver

The block diagram shown in Figure 6.1 includes sensitivity time control (STC) attenuation at the RF input. Alternatively adjustable RF attenuation may be used. Either form provides increased dynamic range above that provided by the analog-to-digital (A/D) converters. RF attenuation is described in more detail in Section 6.6. The STC attenuator is followed by an RF amplifier, often referred to as a low-noise amplifier (LNA). This amplifier provides sufficient gain with a low noise figure to minimize the subsequent degradation of the overall radar noise figure by subsequent components. If sufficient gain is provided in the antenna prior to the receiver, it may be possible to eliminate this gain stage. The RF filter provides rejection of out-of-band interference, including rejection at the RF image frequency. After downconversion to IF, a bandpass filter provides rejection of unwanted signals and sets the receiver analog-processing bandwidth. Additional gain is provided at IF to overcome losses and raise the signal level required for subsequent processing and to set the correct signal level into the A/D converters. An IF limiter provides graceful limiting of large signals that would otherwise overload the A/D converters.

The two dominant methods of digitization, IF sampling and analog I/Q demodulation with baseband A/D conversion, are included for illustration in Figure 6.1, though in general, receivers will not include both techniques. Prior to the availability of affordable digital signal processing, a number of functions, such as monopulse comparison, currently performed in the digital domain, were performed using analog processing within the receiver. Readers interested in the details of these analog processing techniques will find details in the first and second editions of this handbook.[1,2]

All but the simplest of radars require more than one receiver channel. Figure 6.1 shows a single receiver channel that may be replicated any number of times depending on the radar system requirements. Monopulse radars typically include three receiver channels, sum, delta azimuth, and delta elevation channels, used to provide improved angle accuracy. Additionally, many military radar systems include a sidelobe blanker or several sidelobe canceler channels to combat jamming. Since the advent of digital beamforming radar systems, the number of receiver channels required has increased dramatically, with some systems now requiring hundreds of receiver channels. In these multichannel receiver systems, close matching and tracking of gain and phase is required. Receiver channel tracking and equalization are discussed in Section 6.11.

The stable local oscillator (STALO) block provides the local oscillator frequencies for downconversion in the receiver and upconversion in the exciter. For true coherent operation, the STALO is locked to a low frequency reference, shown by the reference oscillator in Figure 6.1 that is used as the basis for all clocks and oscillators such as the coherent local oscillator (COHO) within the receiver and exciter. The clock generator provides clocks to the A/D converters and the direct digital synthesizer and provides the basis for the signals that define the radar transmit and receive intervals.

The direct digital synthesizer in Figure 6.1 is used to generate the transmit waveforms at an IF frequency prior to upconversion to the RF output frequency. Filtering in the exciter is required to reject aliased signals from the direct digital synthesizer and unwanted mixer products. RF gain is typically required to provide a sufficient drive level to the transmitter or phased array antenna.

Almost all modern radar systems use digital signal processing to perform a variety of functions, including pulse compression and the discrimination of desired targets from interference on the basis of velocity or the change in phase from one pulse to the next. Previously, pulse compression was performed using analog processing with dispersive delay lines, typically surface acoustic-wave (SAW) devices. Analog pulse compression has largely been replaced by pulse compression using digital signal processing.

In the case of very wideband waveforms, analog stretch processing (see Section 6.3) may be used to reduce the signal bandwidth before subsequent digital signal processing.

The receiver discussed herein focuses on those functions that provide analog processing and digitization of the individual pulse signals with the minimum of distortion, enabling subsequent digital signal processing to maximize the performance of the radar. The digital signal processing function is not normally considered to be part of the receiver.

6.2 NOISE AND DYNAMIC-RANGE CONSIDERATIONS

Receivers generate internal noise that masks weak signals being received from the radar transmissions. This noise contribution, which can be expressed as either a noise temperature or a noise figure, is one of the fundamental limitations on the radar range.

The noise temperature or noise figure of the radar receiver has been reduced to the point that it no longer represents a dominant influence in choosing between available alternatives. It is a paradox that a noise parameter is usually the first characteristic specified for a radar receiver, yet few radars employ the lowest-noise receiver available because such a choice represents too great a sacrifice in other performance parameters.

Cost is rarely a consideration in rejecting a lower-noise alternative. A reduction in requirements for antenna gain or transmitter power invariably produces cost savings far in excess of any added cost of a lower-noise receiver. Other vital performance characteristics that generally dictate the choice of receiver front end include:

- Dynamic range and susceptibility to overload
- Instantaneous bandwidth and tuning range
- Phase and amplitude stability

A direct compromise must be made between the noise figure and the dynamic range of a receiver. The introduction of an RF amplifier in front of the mixer necessarily involves raising the system noise level at the mixer to make the noise contribution of the mixer itself insignificant. Even if the RF amplifier itself has more than adequate dynamic range, the mixer dynamic range has been compromised, as indicated below:

	Example 1	Example 2	Example 3
Ratio of front-end noise to mixer noise	6 dB	10.0 dB	13.3 dB
Sacrifice in mixer dynamic range	7 dB	10.4 dB	13.5 dB
Degradation of system noise temperature due to mixer noise	1 dB	0.4 dB	0.2 dB

The same considerations apply to the setting of the noise level at the input to the A/D converters. Traditionally, the noise contribution of the A/D converter was considered by the system engineers as a separate contribution to the overall radar system noise, distinct from receiver noise, and was accounted for at the system level. Today, it has become common to include the A/D converter noise as part of the overall receiver noise. Consequently, it is important to understand whether or not the contribution of the A/D converter is included in the specification for the noise figure of a receiver.

In active array antennas, and many conventional antennas, low-noise amplifiers (LNAs) establish the system noise floor prior to the receiver input. The noise from the antenna is usually set well above the receiver noise floor such that the receiver has only a small impact on overall system noise. Again, the trade-off must be performed between system dynamic range and noise figure.

Definitions. Dynamic Range represents the range of signal strength over which the receiver will perform as expected. It requires the specification of a minimum level, typically the noise floor, the maximum level that can be handled with some allowable deviation from the ideal response, and the type of signal to be handled. These parameters are defined through a variety of characteristics as described below.

Modern radars systems increasingly rely solely on linear receiver channels followed by digital signal processing, providing both increased flexibility and near ideal signal-detection characteristics. Previously, a variety of limiting or logarithmic receiver approaches were used to perform various signal-processing functions. These receivers must define an allowable error in their outputs relative to their ideal nonlinear response.

Receivers that include some form of gain control must distinguish between instantaneous dynamic range and the total dynamic range that is achieved as a result of programmed gain variation.

Receiver Input Noise Level. Because many radar systems include low-noise amplifiers prior to the input of the receiver, it is important to understand and specify the noise level at the receiver input. This noise level is set by the antenna noise temperature and its total effective noise gain or loss. The noise level can be specified either as an rms power in a specified bandwidth or as a noise power spectral density.

System Noise. The system noise level is the combined antenna and receiver noise. Typically, the receiver input noise will exceed that of the noise due to the receiver itself, so that the receiver has only a small impact on the system noise temperature or noise figure. Thus, when defining dynamic-range parameters, such as signal-to-noise ratio, it is important to specify whether the noise level being referenced is the receiver noise or total system noise.

Minimum Signal of Interest. Minimum signal definitions such as minimum-detectable-signal or minimum-discernable-signal have been used in the past; however, these definitions have become less common due to the extensive use of digital signal-processing techniques. Digital signal processing of the receiver output allows the detection of signals well below the receiver noise floor and the minimum detectable level depends on the nature of the processing performed.

Signal-to-Noise Ratio (SNR). SNR is the ratio of the signal level to that of the noise. SNR is typically expressed in decibels (dB). The maximum receiver SNR is set by the noise contribution and maximum signal capability of every component in the chain; however, since the limiting technology is often the Analog-to-Digital (A/D) converter, the preceding components and gain structure are often chosen such that the maximum SNR is driven by the performance of the A/D converter. More details of the relationship between A/D converter and receiver SNR are included in Sections 6.10 and 6.11.

Spurious Free Dynamic Range (SFDR). SFDR is the ratio of the maximum signal level to that of largest spurious signal created within the receiver. SFDR is typically expressed in decibels (dB). This parameter is determined by a variety of factors including the mixer intermodulation spurious (described in more detail in Section 6.4), the spurious content of the receiver local oscillators, the performance of the A/D converter, and the many sneak paths that may result in unwanted signals coupling onto the receiver signal path.

Intermodulation Distortion (IMD). Intermodulation distortion is a nonlinear process that results in generation of frequencies that are linear combinations of the fundamental frequencies of the input signals. Second and third order intermodulation are the most commonly specified, and the performance of the receiver is usually specified in terms of two-tone second and third order input intercept points. The intercept point is the extrapolated level at which the power in the intermodulation product equals that of the two fundamental signals.

For input signals at frequencies f_1 and f_2, second order intermodulation distortion produces signals at frequencies: $0, f_1 - f_2, f_1 + f_2$, $2f_1$ and $2f_2$. Third order intermodulation distortion produces signals at frequencies: $2f_1 - f_2$, $2f_2 - f_1$, $2f_1 + f_2, f_1 + 2f_2$, $3f_1$ and $3f_2$. For narrow band signals, only the third order products $2f_1 - f_2$ and $2f_2 - f_1$ fall in band, and consequently, third order distortion is typically the primary concern. The power levels of these third order intermodulation products are given by

$$P_{2f1-f2}(dBm) = 2P_{f1}(dBm) + P_{f2}(dBm) - 2P_{IP}(dBm) \tag{6.1}$$

$$P_{2f2-f1}(dBm) = P_{f1}(dBm) + 2P_{f2}(dBm) - 2P_{IP}(dBm) \tag{6.2}$$

where P_{f1}(dBm) = power of input signal at frequency f_1 in dBm
P_{f2}(dBm) = power of input signal at frequency f_2 in dBm
P_{IP}(dBm) = third order intercept point in dBm

Intermodulation can result in a variety of undesirable effects such as

- Intermodulation of clutter returns causing broadening of clutter doppler width, resulting in the masking of targets
- Unwanted in-band signals due to out-of-band interfering signals, resulting in false targets
- Intermodulation products from in-band signals that cannot be readily cancelled through linear cancellation techniques, resulting in susceptibility to jammers

Intermodulation distortion occurs throughout the receiver chain. Consequently, the receiver will have a significantly different input intercept point, depending on the signal frequency relative to the radio frequency (RF), IF, and video filter bandwidths. It is, therefore, important to distinguish between the requirements for in-band and out-of-band intermodulation distortion as different signals have different effects on the receiver.

Cross-Modulation Distortion. Cross-modulation occurs as a result of third order intermodulation, whereby the amplitude modulation (AM) of one signal, typically an unwanted interference signal in the operating RF band but usually outside the tuned signal bandwidth, is transferred onto the desired signal.

The resultant percent AM modulation, $\%d$, on the desired signal is given by[3]

$$\%d = \%u \frac{4P_U}{P_{\text{IP}} + 2P_U} \tag{6.3}$$

where $\%u$ = percent AM modulation of the unwanted signal
P_U = power of unwanted signal
P_{IP} = third order intercept point

Cross modulation can result in the modulation of clutter and target returns due to large amplitude modulated out-of-band interferences resulting in poor clutter cancellation and poor range sidelobe performance.

1 dB Compression Point. The input 1 dB compression point of a receiver is a measure of the maximum linear signal capability and is defined as the input power level at which the receiver gain is 1 dB less than the small signal linear gain. Receiver gain compression can result from compression in amplifiers, mixers, and other components throughout the receiver chain. Typically, the receiver is designed to provide controlled gain compression through a limiting stage at the final IF as described in Section 6.8.

Analog-to-Digital Converter Full Scale. The A/D converter full scale level determines the maximum level that can be digitized. Receivers typically provide controlled limiting (Section 6.8) to prevent the signal level from exceeding the full scale level of the A/D converter. Practical considerations mean that the hard limit level is typically set 1 dB below full scale to prevent overload as a result of component tolerance variations.

Types of Signals. Various types of signals are of interest in determining dynamic-range requirements: distributed targets, point targets, wideband noise jamming, and narrow band interference. If the radar employs a phase-coded signal, the elements of the receiver preceding the decoder will not restrict the dynamic range of a point target as severely as they will for distributed clutter; the time-bandwidth product of the coded pulse indicates the added dynamic range that the decoder will extract from the point targets. Conversely, if the radar incorporates an excessively wide-bandwidth RF amplifier, its dynamic range may be severely restricted due to wideband noise interference.

When low-noise amplifiers (LNAs) are included in the antenna, prior to forming the receive beams, the antenna sidelobe levels achieved are dependent upon the degree to which gain and phase characteristics are similar in all LNAs. Dynamic range has an exaggerated importance in such configurations because matching nonlinear characteristics is impractical. The effect of strong interference—mountain clutter, other radar pulses, or electronic countermeasures (ECM)—entering through the sidelobes will be exaggerated if it exceeds the dynamic range of the LNAs because sidelobes will be degraded. The LNAs are wideband devices, vulnerable to interference over the entire radar operating band and often outside this band; although off-frequency interference is filtered in subsequent stages of the receiver, strong interference signals can cause clutter returns in the LNA to be distorted, degrading the effectiveness of doppler filtering and creating false alarms. This phenomenon is difficult to isolate as the cause of false alarms in such radars owing to the nonrepetitive character of many

sources of interference. In modern radar architectures that employ digital beamforming, nonlinearity at any stage of the receiver channel will create similar problems.

System calibration techniques and adaptive beamforming techniques can compensate for linear gain and phase deviations; however, as for the case of the LNA nonlinearities described above, compensation for nonlinear characteristics is either impractical or impossible when the cause of the nonlinear distortion is outside the digitized bandwidth.

Evaluation. A thorough evaluation of all elements of the receiver is necessary to prevent unanticipated degradation of noise figure or dynamic range. Inadequate dynamic range makes the radar receiver vulnerable to interference, which can cause saturation or overload, masking or hiding the desired signals. A tabular format for such a computation (a typical example of which is shown in Table 6.1) will permit those components that contribute significant noise or restrict the dynamic range to be quickly identified. "Typical" values are included in the table for purposes of illustration.

TABLE 6.1 Noise and Dynamic-Range Characteristics

	Units	Input	STC Attenuator	Amplifier	Bandpass Filter	Mixer	Bandpass Filter	Amplifier	AGC Attenuator	Limiter	A/D Converter
Component Noise Figure	dB		3.0	5.0	0.5	6.5	5.0	4.0	6.0	14.0	
Component Gain	dB		−3.0	12.0	−5.0	−6.5	−0.2	20.0	−6.0	0.0	
Component Output 3rd Order Intercept	dBm		43.0	32.0	50.0	20.0	50.0	38.0	40.0	30.0	
Component Output 1dB Compression Point	dBm		30.0	18.0	40.0	10.0	40.0	23.0	30.0	−1.0	
Cumulative Gain	dB		−3.0	9.0	8.5	2.0	0.0	20.0	14.0	14.0	
Cumulative Noise Figure	dB		3.00	8.00	8.01	8.33	9.13	9.86	9.88	10.29	
Cumulative Output 3rd Order Intercept	dBm		43.0	32.0	31.4	18.8	16.8	34.3	28.1	25.9	
Cumulative Output 1dB Compression Point	dBm		30.0	18.0	17.5	7.4	5.4	21.0	14.9	−1.1	
Receiver Noise Level	dBm/Hz		−174.0	−157.0	−157.5	−163.7	−164.9	−144.1	−150.1	−149.7	
System Noise Level	dBm/Hz	−149.0	−152.0	−139.0	−140.4	−146.9	−148.9	−128.9	−134.9	−134.9	
Bandwidth	MHz		1000	1000	100	100	10	10	10	10	
A/D SNR in Nyquist BW	dB										70.0
A/D Converter Sample Rate	MHz										100.0
A/D Full Scale Level	dBm	−14.0	−17.0	−5.0	−5.5	−12.0	−14.0	6.0	0.0	0.0	0.0
A/D Noise Level	dBm/Hz										−147.0
System Noise Relative to A/D Noise	dB										12.1
Maximum Point Clutter or Target Level	dBm	−20.0	−23.0	−11.0	−11.5	−18.0	−20.0	0.0	−6.0	−6.0	

6.3 BANDWIDTH CONSIDERATIONS

Definitions. The instantaneous bandwidth of a component is the frequency band over which the component can simultaneously process two or more signals to within a specified accuracy. When the term *instantaneous bandwidth* is used as a radar receiver parameter, it refers to the resulting bandwidth set by the combination of RF, IF, video, and digital filtering that occurs within the receiver.

When the radar receiver employs stretch processing (defined later in this section), the RF processing bandwidth is significantly larger than the IF bandwidth. Consequently, the term *instantaneous bandwidth* can be confusing. Confusion can be avoided by using the terms *RF waveform bandwidth, LO linear FM (chirp) bandwidth*, and *IF processing bandwidth*. The relationship between RF, LO, and IF bandwidths used in stretch processing is explained in more detail later.

The tuning range is the frequency band over which the component may operate without degrading the specified performance. Tuning is typically accomplished by adjusting the local oscillator frequency and adjusting the RF filtering characteristics. The frequency range over which the radar operates is often referred to as the *operating bandwidth*.

Important Characteristics. The environment in which a radar must operate includes many sources of electromagnetic radiation, which can mask the relatively weak returns from its own transmission. The susceptibility to such interference is determined by the ability of the receiver to suppress the interfering frequencies if the sources have narrow bandwidth or to recover quickly if they are more like impulses in character. One must be concerned with the response of the receiver in both frequency and time domains.

Generally, the critical response is determined in the IF portion of the receiver; this will be discussed in Section 6.7. However, one cannot ignore the RF portion of the receiver merely by making it have wide bandwidth. Section 6.2 discussed how excessively wide bandwidth can penalize dynamic range if the interference is wideband noise. Even more likely is an out-of-band source of strong interference (e.g., other radars, TV stations, or microwave communication links) that, if allowed to reach this point, can either overload the mixer or be converted to IF by one of the spurious responses of the mixer.

Ideal mixers in a superheterodyne receiver act as multipliers, producing an output proportional to the product of the two input signals. Except for the effect of nonlinearities and unbalance, these mixers produce only two output frequencies, equal to the sum and the difference of the two input frequencies. The nonlinearities and imbalance of mixers is described in more detail in Section 6.4.

The best radar receiver is one with the narrowest RF instantaneous bandwidth commensurate with the radiated spectrum and hardware limitations and with good frequency and impulse responses. A wide tuning range provides flexibility to escape interference, but if the interference is intentional, as in the case of jamming, a change in RF frequency on a pulse-to-pulse basis may be required using switchable or electronically tuned filters. If the RF filtering is located prior to RF amplification, the filter insertion loss will have a dB for dB impact on the receiver noise figure, another sacrifice in noise temperature to achieve more vital objectives. Yttrium iron garnet (YIG) filters and pin diode switched filters have been used to provide the necessary frequency agility.

Stretch Processing. Stretch processing is a technique frequently used to process wide bandwidth linear FM waveforms. The advantage of this technique is that it allows the effective IF signal bandwidth to be substantially reduced, allowing digitization and subsequent digital signal processing, at more readily achievable sample rates. By applying a suitably matched chirp waveform to the receiver first LO, coincident with the expected time of arrival of the radar return, the resultant IF waveform has a significantly reduced bandwidth for targets over a limited range-window of interest. Provided that the limited-range window can be tolerated, a substantially reduced processing bandwidth allows more economical A/D conversion and subsequent digital signal processing. It also allows a greater dynamic range to be achieved with lower-rate A/D converters than would be achievable if digitization of the entire RF signal bandwidth were performed.

If the LO chirp rate is set equal to the received signal chirp rate of a point target, the resultant output is a constant frequency tone at the output of the stretch processor receiver, with frequency $\Delta tB/T$, where Δt is the difference in time between the received signal and the LO chirp signal, and B/T is the waveform chirp slope (chirp bandwidth/pulse width). Target doppler is maintained through the stretch processing, producing an output frequency offset equal to the doppler frequency, though the wide percentage bandwidth often used means that the doppler frequency can change significantly over the duration of the pulse.

Ignoring the effect of target doppler, the required RF signal bandwidth is equal to the transmitted waveform bandwidth. Given the RF signal bandwidth B_R, the received pulse width T_R, and the range interval ΔT, the required LO reference waveform duration is given by

$$T_L = T_R + \Delta T \tag{6.4}$$

the LO reference chirp waveform bandwidth is given by

$$B_L = \frac{T_R + \Delta T}{T_R} B_R \tag{6.5}$$

and the IF processing bandwidth is given by

$$B_I = \frac{\Delta T}{T_R} B_R \tag{6.6}$$

6.4 RECEIVER FRONT END

Configuration. The radar *front end* consists of a low-noise amplifier (LNA) and bandpass filter followed by a downconverter. The radar frequency is downconverted to an IF, where filters with suitable bandpass characteristics are physically realizable. The mixer itself and the preceding circuits are generally relatively broadband. Tuning of the receiver, between the limits set by the preselector or mixer bandwidth, is accomplished by changing the LO frequency. Occasionally, receivers will include filtering before the LNA in order to limit the effects of intermodulation distortion that can occur in the LNA. Even when filtering is included before the LNA, a second filter is often still required between the LNA and the mixer in order to reject the amplifier noise at the image frequency. Without this filter, the noise contribution of a broadband LNA would be doubled.

The receiver front end may also include a limiter, used to protect the receiver circuitry from damage due to high power that may occur either from leakage during transmit mode or as a result of interference from another system such as a radar at close range. Front-end limiters are discussed in more detail in Section 6.8.

The radar or receiver front end often includes some form of gain or attenuation control as shown in Figure 6.1. Gain control is described in more detail in Section 6.6.

Effect of Characteristics on Performance. Noncoherent pulse radar performance is affected by front-end characteristics in three ways. Noise introduced by the front end increases the radar noise temperature, degrading sensitivity, and limits the maximum range at which targets are detectable. Front-end saturation on strong signals may limit the minimum range of the system or its ability to handle strong interference. Finally, the front-end spurious performance affects the susceptibility to off-frequency interference.

Coherent radar performance is even more affected by spurious mixer characteristics. Range and velocity accuracy is degraded in pulse doppler radars; stationary target cancellation is impaired in MTI (moving-target indication) radars; and range sidelobes are raised in high-resolution pulse compression systems.

Spurious Distortion of Radiated Spectrum. It is a surprise to many radar engineers that components of the radar receiver can cause degradation of the radiated transmitter spectrum, generating harmonics of the carrier frequency or spurious doppler spectra, both of which are often required to be 50 dB or more below the carrier. Harmonics can create interference in other electronic equipment. Spurious doppler spectra levels are dictated by requirements to suppress clutter interference through doppler filtering.

Harmonics are generated by any component that becomes nonlinear when subjected to the power level created by the transmitter and that passes those harmonics to the antenna. Gaseous or diode receiver-protectors are designed to be nonlinear during the transmitted pulse and reflect the incident energy back toward the antenna. Isolators or circulators are often employed to absorb most of the reflected fundamental, but they are generally much less effective at the harmonics. Moreover, these ferrite devices are nonlinear devices and can generate harmonics.

Spurious doppler spectra are created by any process that does not reoccur identically on each transmitted pulse. Gaseous receiver-protectors ionize under transmitter power levels, but there is some small statistical variation in the initiation of ionization on the leading edge of the pulse and in its subsequent development. In radars demanding high clutter suppression (in excess of 50 dB), it has sometimes been found necessary to prevent this variable reflected power from being radiated by use of both a circulator and an isolator in the receive path.

Spurious Response of Mixers. The ideal mixer acts as a multiplier, producing an output proportional to the product of the two input signals. The input RF signal at frequency f_R is frequency shifted or modulated by the LO signal at frequency f_L. Balanced mixers are used to minimize conversion loss and unwanted spurious responses. In active mixers, modulation is performed using transistors, and in passive mixers, the modulation is performed using Schottky-barrier diodes or other solid-state devices (e.g., MESFET) where increased dynamic range is required.

The resulting output signal frequencies ($f_L + f_R$ and $f_L - f_R$) are the sum and difference of the two input frequencies. In practice, all mixers produce unwanted intermodulation

spurious responses[4] with frequencies $\pm nf_L \pm mf_R$ (where m and n are integers), and the degree to which these spurious products impact the radar performance depends upon the type of mixer and the overall radar performance requirements. Analysis of mixer spurious levels is nontrivial, and the receiver designer typically requires tabulated data generated through mixer characterization measurements to predict mixer spurious levels.

Advances in mixer technology have resulted in a wide variety of commercially available devices employing balanced, double balanced, and double-double balanced topologies covering a wide range of RF, LO, and IF frequencies and a range of performance characteristics.

Mixer Spurious-Effects Chart. A graphical display of mixer spurious components up to the sixth order is shown in Figure 6.2. This chart allows identification of those combinations of input frequencies and bandwidths that are free of strong low-order spurious components. Such charts are most useful in determining optimum IF and LO frequencies during the initial design phase. Once the frequency plan has been determined, computer analysis of spurious responses is typically used to ensure spurious free performance over the entire range of LO frequencies and RF and IF bandwidths.

The heavy line in Figure 6.2 represents the desired signal and shows the variation of normalized output frequency $(H - L)/H$ with normalized input frequency L/H. All other lines on the chart represent the unwanted spurious signals. To simplify use of the chart, the higher input frequency is designated by H and the lower input frequency by L.

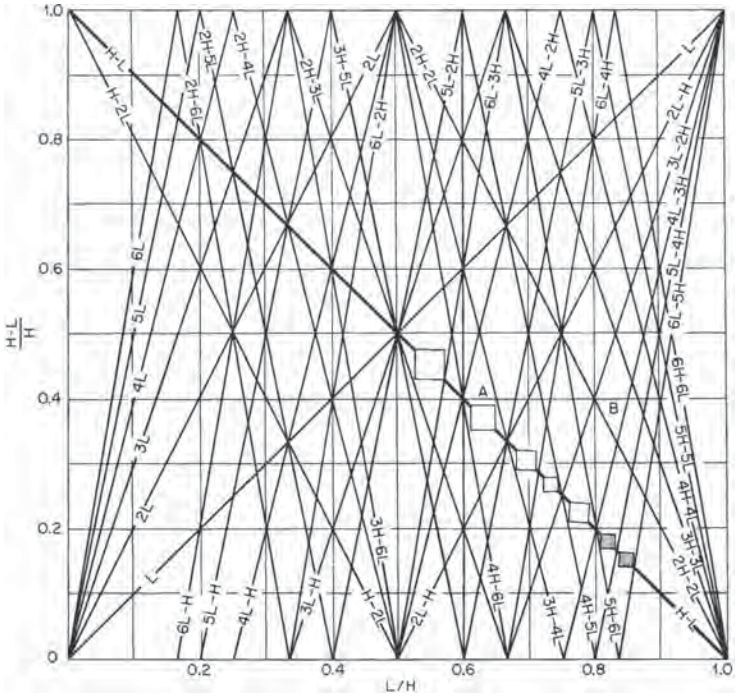


FIGURE 6.2 Downconverter spurious-effects chart: H = high input frequency; L = low input frequency

Seven particularly useful regions have been outlined on the chart. Use of the chart is illustrated by means of the region marked A, which represents the widest available spurious-free bandwidth centered at $L/H = 0.63$. The available RF passband is from 0.61 to 0.65, and the corresponding IF passband is from 0.35 to 0.39. However, spurious IF frequencies of 0.34 ($4H - 6L$) and 0.4 ($3H - 4L$) are generated at the extremes of the RF passband. Any extension of the instantaneous RF bandwidth will produce overlapping IF frequencies, a condition that cannot be corrected by IF filtering. The $4H - 6L$ and $3H - 4L$ spurious frequencies, like all spurious IF frequencies, arise from cubic or higher-order intermodulation.

The available spurious-free bandwidth in any of the designated regions is roughly 10% of the center frequency or $(H - L)/10H$. Thus, receivers requiring a wide bandwidth should use a high IF frequency centered in one of these regions. For IF frequencies below $(H - L)/H = 0.14$, the spurious frequencies originate from high-order terms in the power-series model and are consequently low enough in amplitude that they can often be ignored. For this reason, a low IF generally provides better suppression of spurious responses.

The spurious-effects chart also demonstrates spurious input responses. One of the stronger of these occurs at point B, where the $2H - 2L$ product causes a mixer output in the IF passband with an input frequency at 0.815. All the products of the form $N(H - L)$ produce potentially troublesome spurious responses. These frequencies must be filtered at RF to prevent their reaching the mixer. If sufficient filtering cannot be applied prior to the mixing process, spurious products that fall within the operating band will no longer be filterable, which will seriously degrade system performance.

Spurious responses not predicted by the chart occur when two or more RF input signals produce other frequencies by intermodulation that lie within the RF passband.

Image-Reject Mixer. A conventional mixer has two input responses at points above and below the LO frequency where the frequency separation equals the IF. The unused response, known as the *image*, is suppressed by the image-reject or single-sideband mixer shown in Figure 6.3. The RF hybrid produces a 90° phase differential between the LO inputs to the two mixers. The effect of this phase differential on the IF outputs of the mixers is a +90° shift in one sideband and a −90° shift in the other. The IF hybrid, adding or subtracting another 90° differential, causes the high-sideband signals to add at one output port and to subtract at the other. Where wide bandwidths are involved, the IF hybrid is of the all-pass type. In practice, image reject mixers often do not provide sufficient rejection of the image response alone without filtering. In this case, they can be used in conjunction with an image rejection filter, reducing the magnitude of rejection required by the filter.

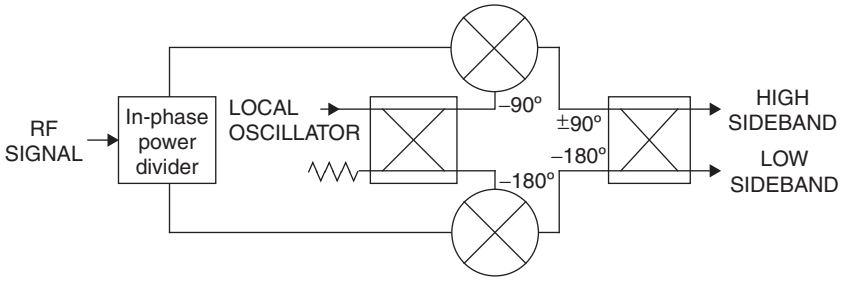


FIGURE 6.3 Image reject mixer

Characteristics of Amplifiers and Mixers. Noise figure, amplifier gain, mixer conversion loss, 1 dB compression point, and third order intercept point are the most common performance parameters specified for amplifiers and mixers. Occasionally, a second order intercept point specification is also required for very wide bandwidth signals. It should be noted that for amplifiers, compression point and third order intercept are usually specified at their output whereas for mixers these parameters are usually specified at their input.

Additional specifications for mixers include LO drive power, port-to-port isolation, and single tone intermodulation levels. The LO drive power specification defines how much LO power is required by the mixer to meet its specified performance levels. Typically, the higher the LO power, the higher the 1 dB compression point and third order intercept point. Radar receivers often require high LO drive level mixers in order to meet the challenging dynamic-range requirements. The port-to-port isolation is used to determine the power level coupled directly between the mixer ports without frequency translation. The single tone intermodulation levels specify the levels of the $\pm nf_L \pm mf_R$ spurious signals, as discussed previously.

6.5 LOCAL OSCILLATORS

Functions of the Local Oscillator. The superheterodyne receiver utilizes one or more local oscillators and mixers to convert the signal to an intermediate frequency that is convenient for filtering and processing operations. The receiver can be tuned by changing the first LO frequency without disturbing the IF section of the receiver. Subsequent shifts in intermediate frequency are often accomplished within the receiver by additional LOs, generally of fixed frequency. These LOs are generally also used in the exciter to upconvert modulated waveforms to RF for output to the transmitter.

In many early radars, the only function of the local oscillators was conversion of the input signal frequency to the correct intermediate frequency. Many modern radar systems, however, coherently process a series of returns from a target. The local oscillators act essentially as a timing standard by which the signal delay is measured to extract range information, accurate to within a small fraction of a wavelength. The processing demands a high degree of phase stability throughout the radar.

STALO Instability. The first local oscillator, generally referred to as a *stable local oscillator (STALO)*, typically has the greatest effect on receiver-exciter stability; however, when evaluating the overall performance, other contributions should not be neglected. Advances in state-of-the-art STALO oscillator performance and the stringent clutter cancellation requirements of modern radars means that the phase noise of all oscillators and timing jitter of A/D converter and D/A converter clocks and T/R strobes may be significant.

The short-term stability requirements of the STALO are generally characterized by device noise relative to carrier (dBc), specified in terms of a phase noise spectrum and measured in the frequency domain. Long-term stability is typically characterized by aging and environmental effects, specified in terms of frequency drift and measured using an Allan Variance[5] technique. Requirements are typically specified in terms of an absolute frequency tolerance or a maximum frequency deviation over some time interval.

It should be noted that measurements of phase noise are typically performed by measurement of double-sideband noise, the sum of the power in both the upper and lower sidebands, but more typically reported and specified as single sideband (SSB) values. Double-sideband noise can be translated to a single-sideband value by subtracting 3 dB. Unequal sideband power can only result from additive signals or noise or correlated amplitude and phase noise components.

Amplitude modulation (AM) of the STALO is typically not a significant factor as it is usually at a lower level than the phase noise (at small offset frequencies from carrier) and can be further reduced through limiting. Modern mixers typically provide a significant reduction in the effect of STALO amplitude modulation as their conversion gain is relatively insensitive to LO power variation when operated at their specified drive level.

For systems requiring high sensitivity, AM noise can become disruptive if unintentional conversion of AM to PM noise occurs in the receiver chain. This process can occur via suboptimum component bias techniques where high amplitude signals or noise create a phase shift resulting in another phase noise contribution to the receiver chain.

Vibration Sensitivity. In addition to the phase noise generated by the STALO in a benign environment, sources of unwanted phase modulation include the effects of power supply ripple and spurious signals as well as mechanical or acoustic vibration from fans, motors, and other sources. The effects of vibration can be severe, especially in airborne environments where high vibration levels are present. The vibration sensitivity of an oscillator is specified by the factional frequency vibration sensitivity, commonly know as the *g-sensitivity*. Typically, a single constant value is specified. In practice, the sensitivity varies significantly with vibration frequency and is different for each axis. Equation 6.7 can be used to determine the effect on oscillator phase noise due to random vibration in each axis.[6]

$$L(f_v) = 20\log_{10}\left[\frac{\Gamma_i f_0 \sqrt{\gamma_i(f_v)}}{f_v}\right] \qquad \text{dBc SSB in a 1 Hz bandwidth} \qquad (6.7)$$

where f_v = vibration frequency (Hz)
f_0 = oscillator frequency (Hz)
Γ_i = oscillator fractional frequency vibration sensitivity (g^{-1}) in axis i
$\gamma_i(f_v)$ = vibration power spectral density (g^2/Hz) in axis i at the vibration frequency f_v

The composite STALO vibration sensitivity (Γ) is defined by the root sum square of the sensitivity in each of the three prime axes, as shown in Eq. 6.8

$$|\Gamma| = \sqrt{\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2} \qquad (6.8)$$

Range Dependence. Most modern radars use the STALO in both the receiver for downconversion and the exciter for upconversion. This double use of the STALO introduces a dependence on range of the clutter and exaggerates the effect of certain unintentional phase-modulation components by 6 dB, the critical frequencies being those which change phase by odd multiples of 180° during the time period between transmission and reception of the clutter return from a specified range.

This range-dependent filter characteristic is given by

$$|F_R(f_m)|^2 = 4\sin^2(2\pi f_m R/c) = 4\sin^2(\pi f_m T) \tag{6.9}$$

where f_m = modulation frequency (Hz)
R = range (m)
c = propagation velocity, 3×10^8 (m/s)
T = time delay = $2R/c$ (s)

A short time delay can tolerate much higher disturbances at low modulation frequencies, as illustrated by the two cases in Figure 6.4. Consequently, the effects of STALO stability need to be computed for several time delays or ranges to ensure sufficient stability exists for the intended application.

Close to carrier phase modulation is typically dominated by that of the oscillators due to the inherent feedback process within the oscillator circuitry. Noise contributors within the oscillator loop that exhibit a $1/f$ characteristic (10 dB/decade) noise slope, are enhanced by 20 dB via the feedback mechanism with a resulting net $1/f^3$ characteristic (30 dB/decade) noise signature close to carrier, within the oscillator loop bandwidth. Outside this loop bandwidth, the oscillator noise signature resumes a $1/f$ slope until reaching a flat thermal noise floor. At larger frequency offsets, significant noise contributions can result from other components such as amplifiers in the STALO signal path. Depending on the location of these amplifiers, they may either create phase modulation that is common to both the receiver and exciter (correlated noise) or add phase noise to only the receiver or exciter (uncorrelated noise). Uncorrelated or uncommon noise is not subject to the range dependent factor described above so it must be accounted for separately. Other significant contributors of uncommon noise are the noise on the exciter waveform before upconversion, along with amplifiers in the receiver and exciter signal paths.

The undesired SSB phase noise after downconversion by the STALO is the sum of the uncommon phase noise and the common phase noise reduced by the range factor.

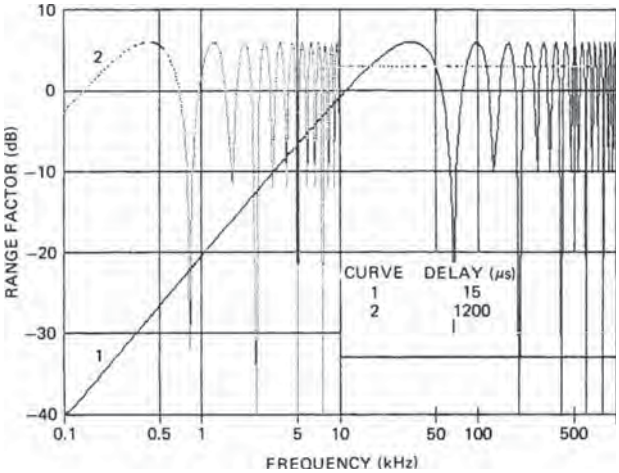


FIGURE 6.4 Effect of range delay on clutter cancellation

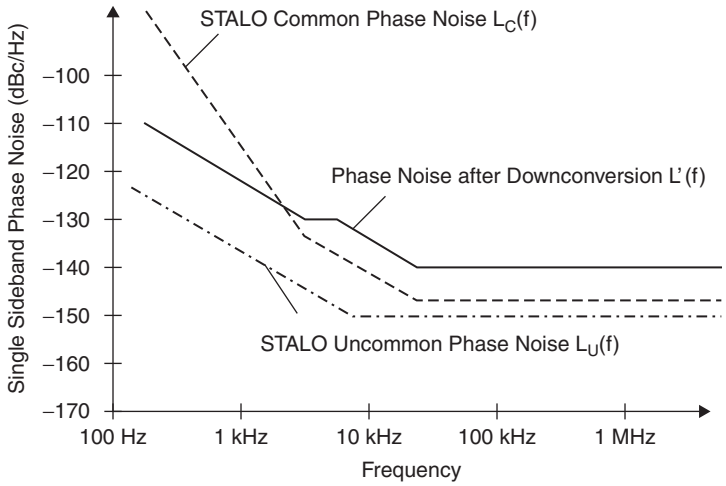


FIGURE 6.5 Phase noise components

Figure 6.5 illustrates typical common and uncommon phase noise components and the resulting mixer output phase noise as calculated using

$$L'(f) = L_C(f) \mid F_R(f) \mid^2 + L_U(f) \tag{6.10}$$

where
$L_C(f)$ = STALO SSB phase noise spectrum common to the receiver and exciter
$L_U(f)$ = total receiver-exciter uncorrelated STALO SSB phase noise
$F_R(f)$ = range dependence factor

Residue Power and MTI Improvement Factor. Subsequent stages of the receiver and signal processor have responses that are functions of the doppler modulation frequency, so the output spectrum can be obtained by combining the responses of these filters with the spectrum present at the mixer input. In MTI systems, it is common to describe the ability to suppress clutter in terms of an MTI improvement factor. The MTI improvement factor I is defined as the signal-to-clutter ratio at the output of the clutter filter divided by the signal-to-clutter ratio at the input of the clutter filter, averaged uniformly over all target radial velocities of interest. The MTI improvement factor limitation due to the STALO may be expressed as the ratio of the STALO power to the total integrated power of the return modulation spectrum it creates at the output of the MTI filters. Figure 6.6 illustrates the effect of the overall filtering, consisting of MTI filtering and receiver filtering on the residue power spectrum.

The integrated residue power due to the STAMO phase noise is given by

$$P_{\text{residue}} = \int_{-\infty}^{\infty} \mid H(f) \mid^2 L'(f) df \tag{6.11}$$

where
$H(f)$ = combined response of receiver and doppler filters, normalized to 0 dB noise gain
$L'(f)$ = phase noise after downconversion as defined in Eq. 6.10

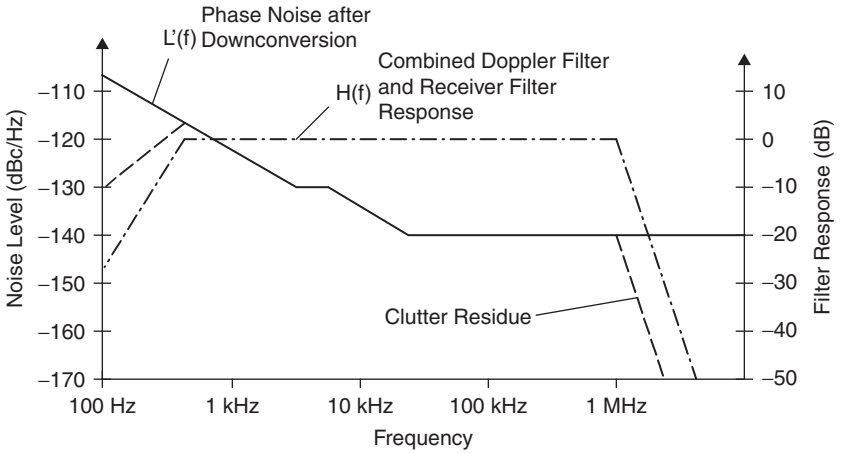


FIGURE 6.6 Clutter residue due to LO phase noise

and the limit on the MTI improvement factor due to the STALO phase noise is given by

$$I = -10 \log_{10} P_{\text{residue}} \tag{6.12}$$

If the radar utilizes more than one doppler filter, the effect of STALO instability should be calculated for each individually.

Pulse Doppler Processing. In pulse doppler systems, a series of pulses are transmitted at a fixed pulse repetition frequency (PRF), and doppler processing is performed within the digital signal processor, using samples separated at the PRF rate. The resulting sampling of the receiver output at the PRF produces aliasing of the phase noise spectrum periodically at the PRF interval, as shown in Figure 6.7, where each curve represents the phase noise at the output of the receiver, including the effects of receiver filtering and offset by a multiple of the PRF frequency. The combined phase noise due to each aliased component is calculated using Eq. 6.13 with the result illustrated in Figure 6.8. This sampled phase noise spectrum provides a method for comparing different LO phase noise profiles and their relative impact on the overall performance of the system.

$$\hat{L}(f) = \sum_{k=-\infty}^{\infty} \left[L'(f + kf_{\text{PRF}}) \, | \, H(f + kf_{\text{PRF}}) \, |^2 \right] \tag{6.13}$$

Sinusoidal Modulations. Radar performance is affected by both random and sinusoidal modulations. Sinusoidal modulations can have a significant impact on radar performance, though the degree to which they cause degradation often depends on their relationship to the radar PRF and their magnitude relative to the random modulations. Examples of such undesired sinusoidal modulations are in-band, unfilterable mixer products, or leakage due to insufficient isolation between signal sources within a receiver or exciter. In addition to external sources of interference, the radar designer

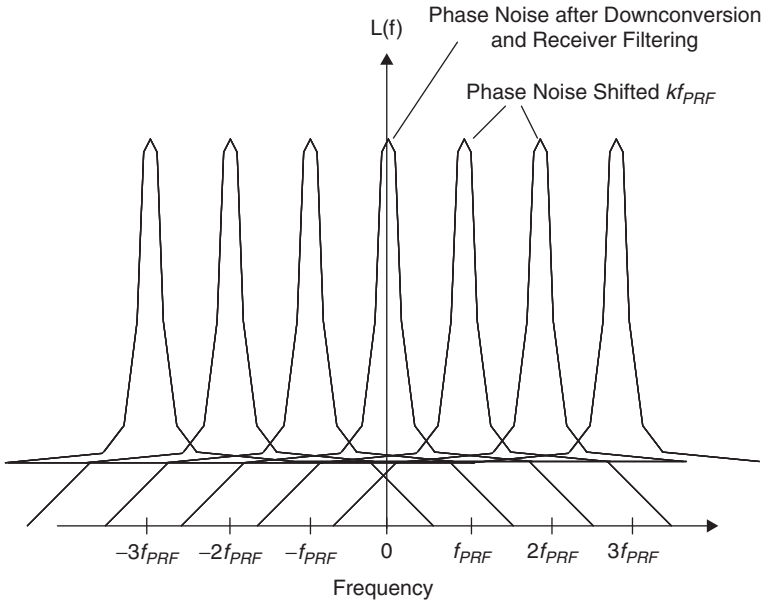


FIGURE 6.7 Phase noise aliasing in a pulse doppler system

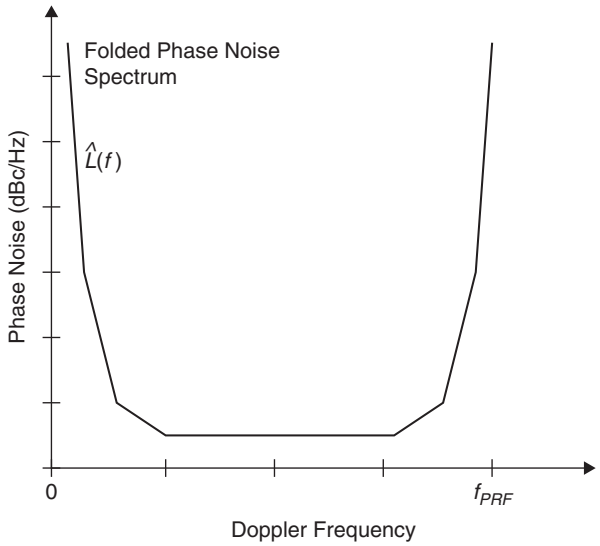


FIGURE 6.8 Sampled phase noise spectrum due to phase noise aliasing

must be concerned with internal signal sources. MTI and pulse doppler radars are particularly susceptible to any such internal oscillators that are not coherent, i.e., that do not have the same phase for each pulse transmission. The effect of the spurious signal is then different for each return, and the ability to reject clutter is degraded.

A truly coherent radar generates all frequencies, including its inter-pulse periods, from a single frequency reference. This fully coherent architecture insures that both the desired frequencies and all the internally generated spurious signals are coherent, eliminating the degradation of clutter rejection.

Many radar systems are pseudo-coherent. The same oscillators are used in both transmit and receive but not necessarily coherent with each other. The result is that the phase of the target remains constant, but the phase of many of the spurious signals varies from pulse to pulse. In this type of configuration, signal isolation and frequency architecture is critical to minimize the occurrence of spurious signals that could erroneously be interpreted as false targets.

COHO and Timing Instability. The majority of this discussion has focused on the STALO as the major contributor to receiver stability. Other contributors such as the second LO, the coherent oscillator (COHO) (if used), A/D and D/A converter clocks can all become significant. A/D and D/A converter clock jitter becomes increasingly significant as sample rates and IF frequencies are increased. The effects of A/D and D/A converter clock phase noise and jitter is described in Sections 6.10 and 6.13. The jitter on timing strobes used to perform transmit/receive (T/R) switching is typically less stringent than that of A/D clocks, as it does not have a direct impact on the signal phase. However, if components such as transmit/receive switches or power amplifiers have a transient phase response of significant duration, time jitter on the switching time can be translated into a phase modulation of the transmitter or receiver signal.

Total Radar Instability. The primary sources of radar instability are usually the receiver-exciter common phase noise, receiver and exciter uncommon phase noise, and the transmitter phase noise. If the spectra of these components are available, either through measurements or through predictions based on similar devices, the convolution of receiver-exciter common phase noise, modified by the range-dependent effect with the other components, provides an estimate of the spectrum of returns from stable clutter, which is then modified by the receiver filters and integrated to obtain the residue power caused by these contributors. These procedures are employed to diagnose the source of radar instability in an existing radar or to predict the performance of a radar in the design stage and to allow the allocation of stability requirements to critical components or subsystems within the radar.

Measurement of total radar instability can be conducted with the radar antenna search-lighting a stable point clutter reflector that produces a signal return close to (but below) the dynamic-range limit of the receiver. Suitable clutter sources are difficult to find at many radar sites, and interruption of rotation of the antenna to conduct such a test may be unacceptable at others; in this case, a microwave delay line can be employed to feed a delayed sample of the transmitter pulse into the receiver. All sources of instability are included in this single measurement except for any contributors outside the delay-line loop. It is important to recognize that timing jitter does not produce equal impact on all parts of the return pulse and generally has minimal effect on the center of the pulse, so it is essential to collect data samples at a multiplicity of points across the return, including leading and trailing edges. The total radar instability is the ratio of the sum of the multiplicity of residue powers at the output of the doppler filter to the sum of the powers at its input, divided by the ratio of receiver noise at these locations. Stability is the inverse of this ratio; both are generally expressed in decibels.

In radars with phase-coded transmission and pulse compression receivers, residue may be significant in the range sidelobe region as well as in the compressed pulse, caused by phase modulation during the long transmitted pulse rather than solely from pulse to pulse. Measurement of stability of such radars must employ a very large number of data points to obtain an answer valid for clutter distributed in range.

In addition to the amplitude and phase noise of the receiver-exciter and the transmitter, mechanically scanning antennas produce a modulation that is predominantly AM. The combined effect is the sum of the residue powers produced by each component individually.

Low Noise Frequency Sources. Many radar systems operate over a range of RF frequencies, requiring a number of LO frequencies that are typically generated using frequency synthesis. Frequency synthesis is the process of creating one or more frequencies from a single reference frequency using frequency multiplication, division, addition, and subtraction to synthesize the required frequencies. The fundamental building block of any frequency synthesis approach is the oscillator. Crystal oscillators have historically been the most common source technology. VHF crystal oscillators employing doubly-rotated (SC, IT, etc.) crystal resonators are able to support higher power levels than single axis crystals. This enables them to achieve lower phase noise and improved vibration immunity due to properties unique to the particular axis of rotation. Frequency multiplication of these VHF sources is often used to generate the radar RF frequencies required; however, this multiplication process results in increase in phase noise performance by $20 \log_{10}(M)$ dB where M is the multiplication factor. A variety of other source technologies, such as Surface Acoustic Wave (SAW) oscillators, have been exploited to achieve improved phase noise performance. SAW oscillators enable lower far-from-carrier phase noise, largely due to their higher frequency operation and the resulting lower frequency multiplication factor required to generate the equivalent radar RF output frequencies.

Very accurate frequency timing is often required in radars where coordination or hand-off from one radar to another, or communication to a missile in flight, is required. This is typically the case where a search radar acquires a target and queues a precision tracking radar. Accurate timing for these applications may be achieved by phase locking the low phase noise radar oscillators to a low frequency reference generated from either a rubidium oscillator or a GPS receiver. In this configuration, the long-term stability of the reference oscillator is superior to that of the radar oscillator, and the short-term stability of the radar oscillator is superior to that of the reference oscillator. The phase lock loop (PLL) architecture is established to exploit the strengths of both technologies by selecting a PLL bandwidth at the offset frequency where the source stabilities cross over. For typical radar and reference oscillator technologies, this usually occurs in the 100 Hz to 1 kHz offset region.

Frequency Synthesis Techniques. The most common techniques are direct synthesis, direct digital synthesis, and frequency multiplication. Direct synthesis is the process of generating frequencies through the multiplication and mixing of a number of signals at different frequencies to produce the required output frequency. Frequency multiplication and direct digital synthesis are described in Section 6.13. Conventional phase locked loop synthesizers are occasionally used, but their frequency switching times and phase settling responses are generally inadequate to meet the stringent radar receiver-exciter requirements. Phase locked loops are more likely used to lock fixed high-frequency oscillators to stable low-frequency references to

ensure coherence of all oscillators within the receiver-exciter and obtain an optimum balance of long- and short-term stability.

Coherence After Frequency Switching. Long range radars often transmit a series of pulses before receiving returns from the first in the sequence. Pulses may be transmitted at a number of different operating frequencies requiring switching of the LO frequency between pulses. If target returns are processed coherently, the phase of the LO signal must be controlled such that each time it switches to a particular frequency, the phase of the LO is the same phase that it would have been had no frequency switching occurred. This requirement drives the architecture used to generate LO frequencies. Generating all the frequencies from a single reference frequency does not guarantee phase coherence when frequency switching occurs. Three sources of phase ambiguity are common: frequency dividers, direct digital synthesizers, and voltage controlled oscillators (VCO). Frequency dividers produce an output signal that can have any one of N phases, where N is the divide ratio; switching dividers can result in phase ambiguity of $2\pi/N$. If frequency dividers are used in the frequency synthesis process, they must be operated constantly without switching the input frequency or divide ratio to avoid this phase ambiguity. Direct digital synthesizers (DDSs) can be used either to generate LO frequencies directly or to generate modulated waveforms prior to upconversion. When pulse-to-pulse phase coherence is required, the starting phase is reset to zero at the start of each pulse. If all the LO frequencies used are multiples of the pulse repletion frequency, the resulting phase will be the same for each pulse. VCOs can be used to create a tunable LO but are usually phase locked to another stable source for improved stability. The tuning voltage design and filter capacitor technology used to achieve phase lock must be carefully designed to ensure rapid voltage and stored charge transitions. Otherwise, the VCO may properly acquire and achieve phase lock, but the residual voltage decay from the transition will manifest itself in an insidious phase ambiguity called *post-tuning drift*.

Stretch Processing. In stretch processing, the LO signal frequency is modulated with a chirp waveform similar to that of the received signal to reduce the bandwidth of the IF signal as described in Section 6.3. The wideband chirp waveform is typically generated by passing a narrower bandwidth linear frequency modulation (LFM) waveform through a frequency multiplier that increases both the operating frequency and bandwidth of the chirp waveform. Frequency multipliers multiply the phase distortion of the input signal and often have significant phase distortion themselves. Distortion of the LO chirp signal phase can have a significant effect on the compressed pulse performance, either distorting the compressed pulse shape or degrading sidelobe performance (Section 6.13). Phase errors can be measured using a test target injected into the receiver and measuring the phase ripple at the receiver output. By performing this measurement with targets injected at different simulated ranges, the errors associated with the receiver LO and test signal can be separated. Correction of receiver LO phase distortion can be readily corrected when using a direct digital synthesizer as described in Section 6.13.

6.6 GAIN CONTROL

Sensitivity Time Control (STC). The search radar detects returns of widely differing amplitudes, often so great that the dynamic range of a fixed-gain receiver will be exceeded. Differences in return strength are caused by differences in radar cross

sections, in meteorological conditions, and in range. The effect of range on radar return strength overshadows the other causes and can be mitigated by a technique known as *sensitivity time control*, which causes the radar receiver sensitivity to vary with time in such a way that the amplified radar return strength is independent of range.

Time sidelobes of compressed pulses in radars that transmit coded waveforms can be degraded by STC. Gradual changes can usually be tolerated, but at close range, the rate of change of attenuation can be very large. Most modern radars that include STC use digital STC control, which can lead to large step sizes at close range unless high digitization rates are used. The phase stability of the STC attenuator is also an important consideration as excessive phase variation as a function of attenuation can have a dramatic impact on range sidelobes.

Clutter Map Automatic Gain Control. In some radars, mountain or urban clutter can create returns that would exceed the dynamic range of the receiver. The spatial area occupied by such clutter is typically a very small fraction of the radar coverage, so clutter map AGC has been used as an alternative to boosting the STC curve. This technique uses a digital map to record the mean amplitude of the clutter in each map cell over many scans and adds receiver attenuation where necessary to keep the clutter returns below the saturation level of the receiver.

Programmable Gain Control. Reduced gain may be desirable in a variety of situations such as high clutter or high interference environments or in short range modes. Fixed attenuation is often preferable to STC or clutter map control. High PRF pulse doppler radars, for example, cannot tolerate STC due to the range ambiguity of targets. Additional attenuation may be programmed either manually via operator control or automatically to increase the receiver's large signal handling capability or to reduce its sensitivity.

Gain Normalization. Receiver gain can vary due to component tolerances, frequency response, variation with temperature, and aging. Accurate receiver gain control is required for a variety of reasons that include target radar cross-section measurement, monopulse angle accuracy, maximizing the receiver dynamic range, and noise level control. Digital gain control permits the calibration of receiver gain by injecting test signals during radar dead time or during some scheduled calibration interval. Calibration coefficients can be stored as a function of commanded attenuation, operating frequency, and temperature as needed. Measurements over time can also be used to assess component aging and potentially predict receiver failure prior to degradation beyond acceptable limits. Accurate gain control is essential for receiver channels used to perform monopulse angle measurements, where amplitudes received in two or more beams simultaneously are compared to accurately determine the target's position in azimuth or elevation. Receiver dynamic range is maximized with accurate gain control as too little gain can result in noise figure degradation and too much gain results in large signals exceeding the A/D converter full-scale or creating unwanted gain compression, intermodulation, or cross modulation distortion.

Automatic Noise-Level Control. Another widely employed use for AGC is to maintain a desired level of receiver noise at the A/D converter. As will be described in Section 6.10, too little noise relative to the quantization increment of the A/D converter causes a loss in sensitivity. Samples of noise are taken at long range, often beyond the instrumented range of the radar or during some scheduled period. If the radar has

RF STC prior to any amplification, it can be set to full attenuation to minimize external interference with minimal (and predictable) effect on system noise temperature. Most radars employ amplifiers prior to STC, so they cannot attenuate external interference without affecting the noise level. The noise level calibration algorithm must be designed to tolerate external interference and returns from rainstorms or mountains at extreme range.

Another concern with amplification prior to STC is that the noise level at the output of the STC attenuator varies with range. At close range, the noise level into the A/D converter may fall below the quantization interval. Also, a constant noise level as a function of range at the receiver output is desirable in order to maintain a constant false alarm rate. Noise injection after the STC attenuator is used to overcome this problem. A noise source and attenuator are often employed at IF to inject additional noise to compensate for the reduced noise after the STC attenuator. Digital control of the noise injection is synchronized with the STC attenuation to provide an effective constant noise level at the A/D converter input.

Gain Control Components. Most modern radars perform gain control digitally. Digital control permits calibration of each attenuation value to determine the difference between the actual attenuation and that commanded by injecting test signals during dead time.

In the past, gain controlled amplifiers were used extensively to control and adjust receiver gain. Recently, this approach has largely been replaced using digital switched or analog (voltage or current) controlled attenuators distributed throughout the receiver chain. Variable attenuators have a number of advantages over variable gain amplifiers; they typically provide broader bandwidths, greater gain control accuracy, greater phase stability, improved dynamic range, and faster switching speed.

The choice between voltage controlled and switched attenuation depends on tradeoffs between performance of a variety of parameters. Switched attenuators generally provide maximum attenuation accuracy, faster switching speed, improved amplitude and phase stability, greater bandwidth, higher dynamic range, and higher power handling capability. Voltage or current controlled attenuators, controlled via a D/A converter, typically provide improved resolution and lower insertion loss.

Gain control attenuators are often incorporated within the receiver at both RF and IF. RF attenuation is used to provide increased dynamic range in the presence of large target returns. By placing the attenuation as close to the front end as possible, large signals can be handled by minimizing gain compression, intermodulation, or cross-modulation distortion in the majority of receiver components. The disadvantage of using front-end attenuation is that it will typically have a larger impact on receiver noise figure than attenuation placed later in the receiver. This is not usually an issue when the intent of adding attenuation is to desensitize the receiver as is the case for STC. Back-end or IF attenuation is often used to adjust the gain of the receiver to compensate for receiver gain variations due to component variations where receiver noise figure degradation cannot be tolerated.

6.7 FILTERING

Filtering of the Entire Radar System. Filtering provides the principal means by which the radar discriminates between target returns and interference of many types. The filtering is performed by a variety of filters throughout the receiver and

in the subsequent digital signal processing. Most radars transmit multiple pulses at a target before the antenna beam is moved to a different direction, and the multiple returns are combined in some fashion. The returns may be combined using coherent integration or various doppler processing techniques (including MTI) to separate desired targets from clutter. From the radar system standpoint, these are all filtering functions, and in modern radar systems, these functions are performed using digital signal processing on the receiver output I and Q data. These functions are discussed in other chapters of this handbook. The purpose of the filtering within the receiver is to reject out-of-band interference and digitize the received signal with the minimum of error so that optimum filtering can be performed using digital signal processing.

Matched Filtering. Although matched filtering is typically now performed within the digital signal-processing function, the concept is explained here for completeness. The overall filter response of the system is chosen to maximize the radar performance. If the signal spectrum $X(\omega)$, in the presence of white noise with power spectral density $N_0/2$, is processed with a filter with frequency response $H(\omega)$, the resulting signal-to-noise ratio (SNR) at time T is given by[7]

$$\chi = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) H(\omega) e^{j\omega T} d\omega \right|^2}{\frac{N_0}{4\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 \, d\omega} \tag{6.14}$$

The ideal filter response from the standpoint of maximizing SNR is the matched filter that maximizes the SNR at time T_M when

$$H_M(\omega) = X^*(\omega) e^{-j\omega T_M} \tag{6.15}$$

Deviations from the ideal matched filter response $H_M(\omega)$ produce a reduction in SNR termed *mismatch loss*. This loss can occur for a number of reasons such as target doppler or because a filter response is chosen that is different from the matched filter response in order to minimize another parameter such as range sidelobes.

Receiver filtering is often modified for different waveforms used. When radar systems use waveforms of widely varying bandwidths, different I/Q data rates may be used to minimize the digital signal-processing throughput requirements. With different data rates comes the need to adjust the receiver filtering in order to avoid aliasing signals beyond the Nyquist rate. Although these radars adjust their filtering to the waveform bandwidth, they do not typically implement the matched filtering within the receiver. This function is usually implemented in digital signal processing.

Receiver Filtering. Filtering is required at various points throughout the receiver chain including RF, IF, baseband if used, digital filtering prior to decimation (reduction of the sample rate), and as an integral part of I/Q generation.

Section 6.4 described how spurious responses are generated in the mixing process. Unwanted interference signals can be translated to the desired intermediate frequency even though they are well separated from the signal frequency at the input to the mixer. The ability of the radar to suppress such unwanted interference is dependent upon the filtering preceding the mixer as well as on the quality of the mixer itself.

The primary function of RF filtering is the rejection of the image response due to the first downconversion. Image rejection filtering can be alleviated using an image reject mixer; however, the maximum rejection achievable by image reject mixers is typically inadequate without the use of additional rejection through filtering. This image-suppression problem is the reason why some receivers do not translate from the received signal frequency directly to the final intermediate frequency in a single step.

The other spurious products of a mixer generally become more serious if the ratio of input to output frequencies of the downconverter is less than 10. The spurious-effects chart (Figure 6.2) shows that there are certain choices of frequency ratio that provide spurious-free frequency bands, approximately 10% of the intermediate frequency in width. By the use of a high first IF, one can eliminate the image problem and provide a wide tuning band free of spurious effects. Filtering prior to the mixer remains important, however, because the neighboring spurious responses are of relatively low order and may produce strong outputs from the mixer. RF filtering is also important as it reduces out-of-band interference before it can cause intermodulation or cross-modulation distortion within the receiver.

If the receiver operating bandwidth is a large percentage of the RF frequency, some form of switched or tunable RF filtering may be required so that the image response is rejected as it moves through the operating bandwidth. The choice between using switched or tunable filtering depends on the switching speed, linearity, and stability requirements of the receiver. Switched filters provide the fastest response time, with excellent linearity and stability but can be bulky and suffer from the additional loss of the switch components.

An alternate approach that is sometimes used with large operating bandwidths is to first upconvert the input RF signal to an IF frequency higher than the RF operating band. This process virtually eliminates the image response problem, allowing the use of a single RF filter spanning the entire operating bandwidth. Narrow bandwidth filtering can be used on the high IF as defined by the signal bandwidth before downconversion to a lower IF for digitization or baseband conversion.

IF filtering is the primary filtering used to define the receiver bandwidth prior to A/D conversion in receivers using either IF sampling or baseband conversion. In IF sampling receivers, the IF filter acts as the anti-aliasing filter and limits the bandwidth of signals entering the A/D converter. In receivers using baseband conversion, the IF filter sets the receiver bandwidth. Subsequent video filtering should be of greater bandwidth to prevent the introduction of I/Q imbalance due to filter differences between I and Q channels.

In IF sampling receivers, digital filtering is usually the primary means of setting the final receiver bandwidth and provides anti-alias rejection required to prevent aliasing in the decimation of the I/Q data rate. Digital filtering can be precisely controlled, tailored to almost any desired passband and stop band rejection requirements. The digital filters used are typically linear phase FIR filters, but they can also be tailored to compensate for variations in the passband phase and amplitude responses of RF and IF analog filters.

Filter Characteristics. Filter responses are characterized fully by either their frequency response $H(\omega)$ or their impulse response $h(t)$; however, they are usually specified by a variety of parameters as described below. Digital filters may be specified using the same measures, or because they can be specified exactly, they are frequently specified by their transfer function $H(z)$ or impulse response $h(n)$.

Key passband characteristics are: insertion loss, bandwidth, passband amplitude and phase ripple, and group delay. Bandwidths are frequently specified in terms of a 3 dB bandwidth; however, if a low passband variation is required, the specified

bandwidth may be, for example, specified as a 0.5 dB or 0.1 dB bandwidth. Passband amplitude variation relative to the insertion loss is a key parameter that has potential impact on range sidelobes and channel-to-channel tracking. Phase ripple, if specified, is relative to a best-fit linear phase and has similar effects as amplitude ripple. Group delay, the rate of change of phase vs. frequency, is ideally constant for linear phase filters. The absolute value of group delay does not impact the range sidelobe performance; however, the relative group delay between channels must be tightly controlled or compensated in monopulse, sidelobe canceler, and digital beamforming systems.

Although stopband rejection is clearly a key parameter, filters with fast roll-off may not provide the required phase and impulse response characteristics. Figure 6.9 shows the magnitude response of six different fifth order low-pass filters with equal 3 dB bandwidth.[8] The Chebshev filters (0.1 and 0.01 dB ripple) have flat passband response and improved stopband rejection relative to the remaining filters; however, as shown in Figure 6.10 and Figure 6.11, they have inferior phase (group delay) and impulse response characteristics.

Digital filters can be either Finite Impulse Response (FIR) or Infinite Impulse Response (IIR). FIR filters are typically preferred as their finite response is desirable along with their linear phase characteristic. Phase linearity is achieved with the symmetric impulse response condition[9] defined by Eq. 6.16 or the anti-symmetric impulse response conditions defined by Eq. 6.17:

$$h(n) = h(M-1-n) \qquad n = 0, 1, \ldots, M-1 \tag{6.16}$$

where M is the length of the FIR filter impulse response

$$h(n) = -h(M-1-n) \qquad n = 0, 1, \ldots, M-1 \tag{6.17}$$

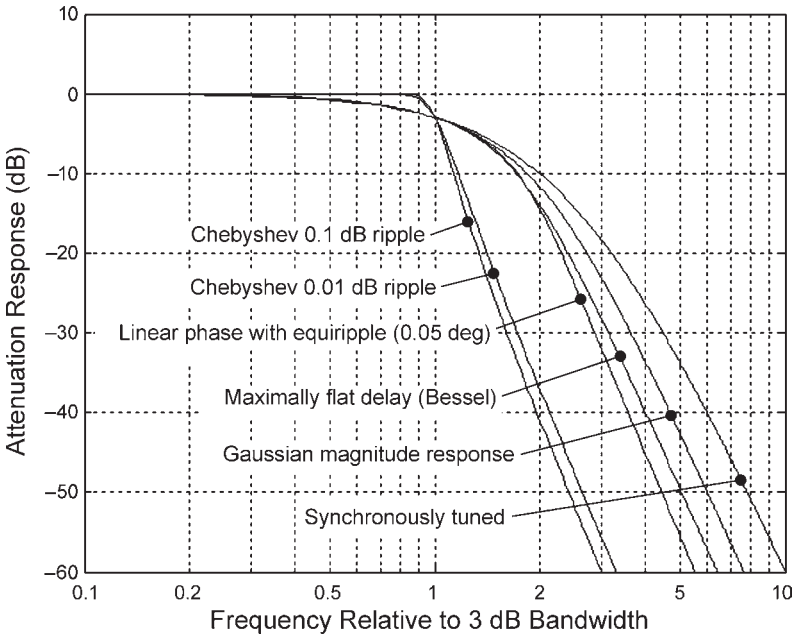


FIGURE 6.9 Magnitude response of lowpass filters

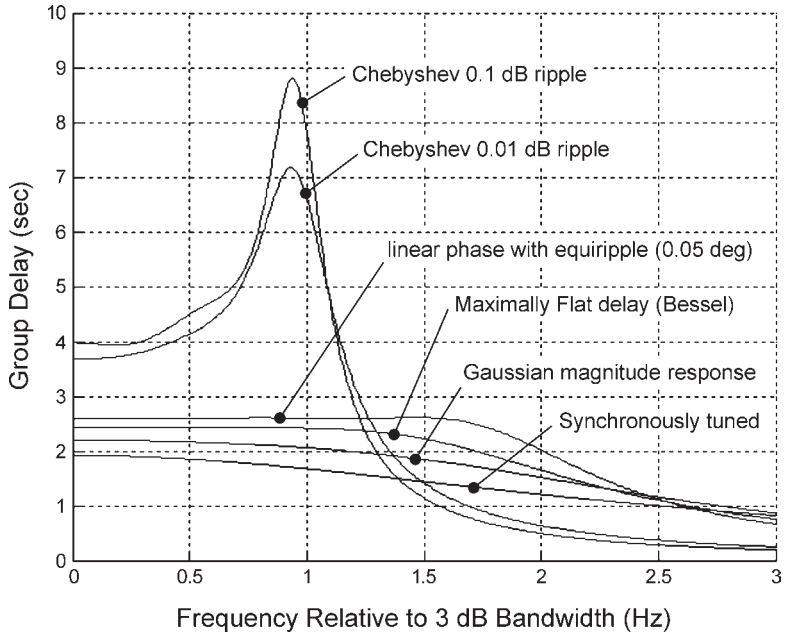


FIGURE 6.10 Group delay response of lowpass filters

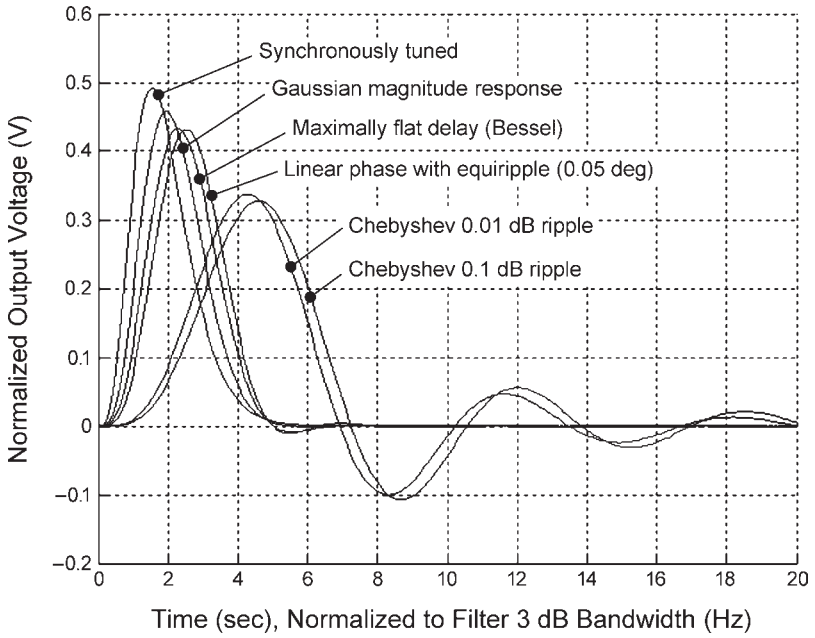


FIGURE 6.11 Normalized impulse response of lowpass filters

Range Sidelobes. Errors in filter responses can produce degradation in pulse compression range sidelobes. The effect of a filter response on range or time sidelobes can be seen by taking the filter impulse response $h(t)$ and adding to this a delayed impulse response $20\log_{10}(\alpha)$ dB below the main response to produce the modified response $h'(t)$, which is given by

$$h'(t) = h(t) + \alpha h(t - T_0) \tag{6.18}$$

Using the property of time shifting of the Fourier transform, the resultant frequency response is given by

$$H'(\omega) = H(\omega) + \alpha e^{-j\omega T_0} H(\omega) \tag{6.19}$$

Thus, for small values of α, the resulting magnitude and phase response is that of the original filter modified by a sinusoidal phase and amplitude modulation as given here:

$$|H'(\omega)| = |H(\omega)|(1 + \alpha\cos(\omega T_0)) \tag{6.20}$$

$$\angle H'(\omega) = \angle H(\omega) - \alpha\sin(\omega T_0) \tag{6.21}$$

Therefore, if there are n ripples across the filter bandwidth B, the range sidelobe occurs at time T_0 given by

$$T_0 = n/B \tag{6.22}$$

Assuming a compressed pulse width of 1/B, values of $n < 1$ will put the range sidelobe within the main lobe of the target return, resulting in a distortion of the mainlobe response.

Channel Matching Requirements. Radar receivers with more than one receiver channel typically require some degree of phase and amplitude matching or tracking between channels. In order to operate effectively, sidelobe canceler channels must track very closely. Constant offsets in gain or phase do not degrade sidelobe canceler performance, but small variations in phase and amplitude across the bandwidth cause significant degradation. For example, achieving a cancellation ratio of 40 dB requires a gain tracking of less than 0.1 dB across the receiver bandwidth. Filters are the main source of amplitude and phase ripple across the signal bandwidth as other components such as amplifiers and mixers are typically relatively broadband. The degree of tracking required for sidelobe canceler operation was previously achieved by providing matched sets of filters with tightly tracking amplitude and phase responses. Modern digital signal processing allows the correction of these channel-to-channel variations using FIR equalization (Section 6.11) or correction in the frequency domain in the digital signal processor, allowing the use of less tightly controlled filters.

6.8 LIMITERS

Applications. Limiters are used to protect the receiver from damage and to control saturation that may occur within the receiver. When received signals saturate some stage of the radar receiver that is not expressly designed to cope with such a

situation, the distortions can result in severely degraded radar performance, and the distortion of operating conditions can persist for some time after the signal disappears. Video stages are most vulnerable and take longer to recover than IF stages; so it is customary to include a limiter in the last IF stage, designed to quickly regain normal operating conditions immediately following the disappearance of a limiting signal. Limiting prior to the A/D converter also prevents the distortion that occurs when signals exceed full-scale. Although A/D converters can often handle modest overload with fast recovery, the distortion that occurs degrades signal processing such as digital pulse compression and clutter rejection. With IF limiting, these harmonics are filtered out using bandpass filtering after limiting prior to A/D conversion, minimizing the degradation due to limiting.

All radar systems contain some form of Transmit/Receive (T/R) device to protect the receive electronics from the high-power transmit signal. In many systems, an RF front-end limiter is also required in order to prevent the receiver from being damaged by high input power levels from the antenna that may occur as a result of leakage from the T/R device during transmit mode or from interference due to jammers or other radar systems. These limiters are typically designed to limit well above the maximum signals to be processed by the receiver.

In the past, limiters were used to perform a variety of analog signal-processing functions. Hard limiters with as much as 80 dB of limiting range were used with some designed to limit on-receiver noise. Applications that utilize hard limiting, including phase-detectors and phase-monopulse receivers, are described in Section 3.10 of the second edition[1] of this handbook. Modern radar systems are mostly designed to maximize the linear operating region, with limiters used only to handle excessively large signals that inevitably exist under worst case conditions.

Characteristics. The ideal limiter is perfectly linear up to the power level at which limiting begins followed by a transition region beyond which the output power remains constant. In addition, the insertion phase is constant for all input power levels, and recovery from limiting is instantaneous. The output waveform from a bandpass limiter is sinusoidal, whereas the output waveform from a broadband limiter approaches a square wave. Deviations from the ideal characteristics can degrade radar performance in a variety of ways.

Linearity Below Limiting. One major drawback of adding a limiter stage to a receiver channel is that it is inherently nonlinear. Since any practical limiter has a gradual transition into limiting, the limiter is often the largest contributor to receiver channel nonlinearity in the linear operating region and can cause significant intermodulation distortion of in-band signals. For this reason, the primary limiting stage is usually located at the final IF stage where maximum filtering of out-of-band interference has been achieved. The lower operating frequency also allows implementation of a limiter that more closely matches the ideal characteristics.

Limiting Amplitude Uniformity. No single-stage limiter will exhibit a constant output over a wide range of input signal amplitudes. One cause is apparent if one considers the effect of a single-stage limiter having a perfectly symmetrical clipping at voltages $\pm E$. For a sinusoidal input, the output signal at the threshold of limiting is

$$v_0 = E\sin(\omega t) \tag{6.23}$$

and when the limiter is fully saturated and the output waveform is rectangular, it is given by the Fourier series:

$$v_o' = \frac{4E}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{1}{n} \sin(n\omega t) \tag{6.24}$$

which is an increase of 20 log $(4/\pi) = 2.1$ dB in the power of the fundamental.

In practice, the amplitude performance is also degraded by capacitive coupling between input and output of each limiting stage, charge storage in transistors and diodes, and *RC* time constants that permit changes in bias with signal level. For these reasons, two or more limiter stages may be cascaded when good amplitude uniformity is required over a wide dynamic range.

Phase Uniformity. The change of insertion phase of the limiter with amplitude is less of a concern for modern radar systems that operate primarily in the linear operating region. However, maintaining constant insertion phase during limiting preserves the phase of target returns in the presence of limiting clutter or interference. The change of insertion phase with signal amplitude is generally directly proportional to the frequency at which it is operated.

Recovery Time. The recovery time of a limiter is a measure of how quickly the limiter returns to linear operation after the limiting signal is removed. Fast recovery is particularly important when the radar is exposed to impulsive interference.

6.9 I/Q DEMODULATORS

Applications. The I/Q demodulator, also referred to as a quadrature channel receiver, quadrature detector, synchronous detector, or coherent detector, performs frequency conversion of signals at the IF frequency to a complex representation, $I + jQ$ centered at zero frequency. The baseband in-phase (I) and quadrature-phase (Q) signals are digitized using a pair of A/D converters providing a representation of the IF signal, including phase and amplitude without loss of information. The resulting digital data can then be processed using a wide variety of digital signal-processing algorithms, depending on the type of radar and mode of operation. Processing such as pulse compression, doppler processing, and monopulse comparison, all require amplitude and phase information. The predominance of digital signal processing in modern radar systems has led to almost universal need for Nyquist rate sampled data. In many modern radar systems, digital I and Q data is now generated using IF sampling followed by digital signal processing used to perform the baseband conversion as described in Sections 6.10 and 6.11. I/Q demodulators are still used, though their use is increasingly limited to wider bandwidth systems where A/D converters are not yet available with the required combination of bandwidth and dynamic range to perform IF sampling.

Implementation. Figure 6.12 shows the basic block diagram of a I/Q demodulator. The IF signal described by Eq. 6.25 is split and fed to a pair of mixers or analog multipliers. The mixer LO ports are fed with a pair of signals in quadrature, generated from the reference frequency signal, or coherent oscillator (COHO), and represented

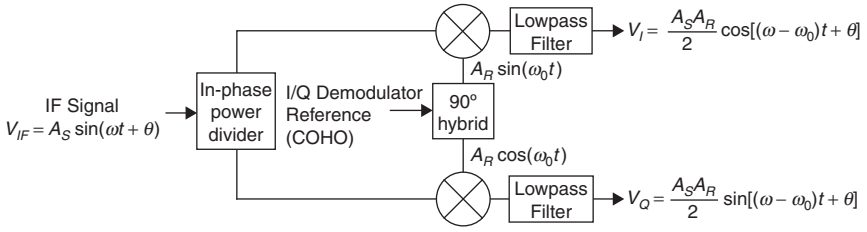

Lowpass Filter
$V_I = \frac{A_S A_R}{2} \cos[(\omega - \omega_0)t + \theta]$
$A_R \sin(\omega_0 t)$
IF Signal
$V_{IF} = A_S \sin(\omega t + \theta)$
In-phase power divider
I/Q Demodulator Reference (COHO)
90° hybrid
$A_R \cos(\omega_0 t)$
Lowpass Filter
$V_Q = \frac{A_S A_R}{2} \sin[(\omega - \omega_0)t + \theta]$

FIGURE 6.12 I/Q demodulator

in complex form in Eq. 6.26. Ignoring any mixer insertion loss or loss associated with the IF split, the complex representation of the mixer output is given by Eq. 6.27. Ideal low-pass filtering rejects the second (sum frequency) term of Eq. 6.27, producing the I/Q demodulator output as represented by Eq. 6.28.

$$V_{\text{IF}} = A_S \sin(\omega t + \theta) = \frac{A_S}{2j}(e^{j(\omega t+\theta)} - e^{-j(\omega t+\theta)}) \tag{6.25}$$

$$V_{\text{COHO}} = A_R\left[\sin(\omega_0 t) + j\cos(\omega_0 t)\right] = jA_R e^{-j\omega_0 t} \tag{6.26}$$

$$V_{\text{IF}}V_{\text{COHO}} = \frac{A_S}{2}(e^{j(\omega t+\theta)} - e^{-j(\omega t+\theta)})A_R e^{-j\omega_0 t} = \frac{A_S A_R}{2} e^{j[(\omega-\omega_0)t+\theta]} - \frac{A_S A_R}{2} e^{-j[(\omega+\omega_0)t+\theta]} \tag{6.27}$$

$$V_I + jV_Q = \frac{A_S A_R}{2}\cos[(\omega - \omega_0)t + \theta] + j\frac{A_S A_R}{2}\sin[(\omega - \omega_0)t + \theta] = \frac{A_S A_R}{2} e^{j[(\omega-\omega_0)t+\theta]} \tag{6.28}$$

In implementing an I/Q demodulator, it is important to provide well-balanced *I* and *Q* channels in order to maximize image rejection, as explained below. The mixers must have DC coupled IF output ports and be presented with a good match at both the wanted low frequency output and the unwanted sum frequency. A match at the sum frequency can be provided using a diplexer filter. Video filtering is required to reject the sum frequency mixer outputs and also provides rejection of wideband noise from the video amplifiers, which would otherwise alias to baseband through the A/D converter sampling process, producing an unwanted degradation of receiver noise figure. Video amplification is often required to increase the signal level to the full-scale signal level of the A/D converter and also allows for impedance matching of the mixer and A/D converter.

The convention for the *I* and *Q* relationship is that the *I* signal phase leads the *Q* signal phase for radar signals with positive doppler (approaching targets). Frequency conversions within the receiver using LO frequencies greater than the RF frequency will cause a doppler frequency inversion, so each conversion must be considered in order to achieve the correct sense of *I* and *Q* at the receiver output. Fortunately, an incorrect *I* and *Q* relationship can easily be fixed either in the receiver or the signal processor, by switching the *I* and *Q* digital data or by changing the sign of either *I* or *Q*.

Gain or Phase Imbalance. If the gains of the *I* and *Q* channels are not exactly equal or if their COHO phase references are not exactly 90 degrees apart, an input signal at frequency ω will create an output at both the desired frequency ω–ω_0 and

at the image frequency $-(\omega - \omega_0)$. The image signals generated by gain and phase imbalance are given by Eq. 6.29 and Eq. 6.30. For small errors, if the ratio of voltage gains is $(1 \pm \Delta)$ or if the phase references differ by $(\pi/2 \pm \Delta)$ radians, the ratio of the spurious image at $-\omega_d$ to the desired output of ω_d is $\Delta/2$ in voltage, $\Delta^2/4$ in power, or 20 (log Δ) −6 in decibels.

$$V_I + jV_Q = E\cos(\omega_d t) + j(1+\Delta)E\sin(\omega_d t) = \left(1 + \frac{\Delta}{2}\right)Ee^{j\omega_d t} - \frac{\Delta}{2}Ee^{-j\omega_d t} \tag{6.29}$$

$$V_I + jV_Q = E\cos(\omega_d t) + jE\sin(\omega_d t + \Delta) = \cos\left(\frac{\Delta}{2}\right)Ee^{j\left(\omega_d t + \frac{\Delta}{2}\right)} - \sin\left(\frac{\Delta}{2}\right)Ee^{-j\left(\omega_d t + \frac{\Delta+\pi}{2}\right)} \tag{6.30}$$

Historically, *I* and *Q* phase and gain corrections have been performed using adjustments in the analog signal paths, as shown in Figure 6.13. Gain errors may be corrected by a change in gain in the IF or video stages of either or both *I* and *Q* channels. Video gain control must be implemented with care as it can exaggerate the nonlinearity of those stages. These corrections can now be implemented more precisely in the digital domain.

A measurement of the signal spectrum at the center of the IF bandwidth indicates the degree of gain and phase imbalance compensation. However, as the following discussion will explain, the suppression of image energy across the IF bandwidth may be substantially less than indicated by this measurement at IF center.

Time Delay and Frequency Response Imbalance. If the responses of the *I* and *Q* channels are not identical across the entire signal bandwidth, unwanted image responses will occur that are frequency dependent. Optimum bandpass filtering should

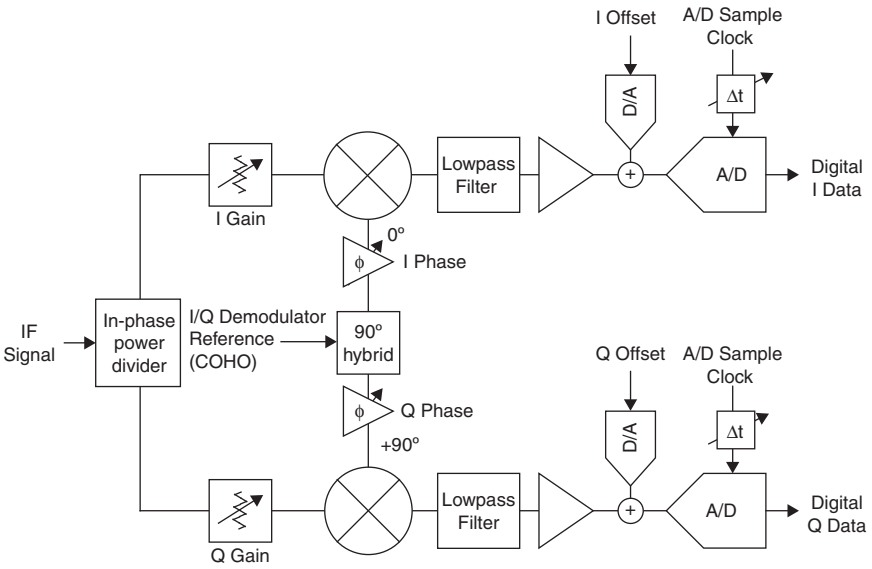


FIGURE 6.13 I/Q demodulator with gain, phase, DC offset, and time-delay adjustments

be at IF, where it affects I and Q channels identically, not at baseband. Video filter bandwidth should be more than half the IF bandwidth and controlled by precision components in order to minimize the creation of image signals. Substituting $\Delta(\omega)$ for Δ in Eq. 6.29 and Eq. 6.30 gives the image components for frequency dependent gain and phase errors. Similarly, substituting $\omega\Delta T$ for Δ in Eq. 6.30 gives the image component due to time-delay imbalance in the I and Q paths. Small time-delay imbalances can be corrected by adding time delay to the A/D sample clock, as shown in Figure 6.13. Large time-delay corrections should be avoided as they can cause problems aligning the I and Q digital data. When adding time delay to the sample clock, care must be taken to avoid adding jitter, which could degrade the A/D converter SNR performance. Time-delay correction can also be implemented effectively in the digital domain, and if frequency dependent phase and amplitude imbalance correction is required, this is most easily and effectively performed in the digital domain using FIR filtering of the I and Q data or by performing corrections in the frequency domain data as part of the radar signal processing.

Nonlinearity in *I* and *Q* Channels. Component tolerances often lead to somewhat different nonlinearities in I and Q, which can generate the variety of spurious doppler components.

The ideal input signal is

$$V = Ae^{j\omega_d t} = I + jQ \tag{6.31}$$

Each video channel response can be expressed as a power series. For simplicity, only symmetrical distortion will be considered. The A/D output, including a residual gain imbalance of Δ, is

$$V'_{\mathrm{IQ}} = V'_I + jV'_Q \tag{6.32}$$

$$V'_I = V_I - aV_I^3 - cV_I^5 \tag{6.33}$$

$$V'_Q = (1 + \Delta)V_Q - bV_Q^3 - dV_Q^5 \tag{6.34}$$

Substitution of Eqs. 6.33 and 6.34 into Eq. 6.31 yields the amplitudes of the spectral components listed in Table 6.2. Note that if the nonlinearities in I and Q were identical ($a = b$; $c = d$), spurious components at -5ω and $+3\omega$ would not be present and the image ($-\omega$) would be proportional to input signal amplitude. Spurious at zero doppler is not due to dc offset; it is the result of even-order nonlinearities that were omitted from the above equations. The negative third harmonic is the dominant component produced by nonlinearity.

TABLE 6.2 Spurious Signal Components Generated by *I*/*Q* Nonlinearity

Signal Frequency	Amplitude of Spectral Component
-5ω	$A^5(c - d)/32$
-3ω	$A^3(a + b)/8 + 5A^5\ (c + d)/32$
$-\omega$	$A(\Delta/2) + 3A^2\ (a - b)/8 + 5A^5\ (c - d)/16$
(Input)ω	$A(1 + \Delta/2) - 3A^2(a + b)/8 - 5A^5(c + d)/16$
$+3\omega$	$A^3(a - b)/8 + 5A^5\ (c - d)/32$
$+5\omega$	$A^5(c + d)/32$

DC Offset. Small signals and receiver noise can be distorted by an offset in the mean value of the A/D converter output unless the doppler filter suppresses this component.

False-alarm control in receivers without doppler filters is sometimes degraded by errors of a small fraction of the least significant bit (LSB), so correction is preferably applied at the analog input to the A/D. DC offsets can be measured using digital processing of the A/D converter outputs and a correction applied using D/A converters, as shown in Figure 6.13. DC offset correction can also be performed effectively in the digital domain, provided that the DC offset at the input of the A/D converter is not so large that it results in a significant loss of available dynamic range.

Many of the I/Q demodulator errors described above are either reduced dramatically or eliminated using IF sampling. This, along with the reduction of hardware required, are the reasons that IF sampling (described in Sections 6.10 and 6.11) is becoming the dominant approach.

6.10 ANALOG-TO-DIGITAL CONVERTERS

The high-speed A/D converter is a key component in receivers of modern radar systems. The extensive use of digital signal processing of radar data has resulted in a demand for converters with both state-of-the-art sampling rates and dynamic range.

Analog to digital converters transform continuous time analog signals into discrete time digital signals. The process includes both sampling in the time domain, converting from continuous time to discrete time signals and quantization, converting from continuous analog voltages to discrete fixed-length digital words. Both the sampling and quantization process produce errors that must be minimized in order to limit the radar performance degradation. In addition, a variety of other errors such as additive noise, sampling jitter, and deviation from the ideal quantization, result in non-ideal A/D conversion.

Applications. The conventional approach of using a pair of converters to digitize the I and Q outputs of an I/Q demodulator is, in many cases, being replaced by digital receiver architectures where a single A/D converter is followed by digital signal processing to generate I and Q data. Digital receiver techniques are described in Section 6.11.

Although the dividing line is arbitrary and advancing with the state-of-the-art, radar receivers are often classified as either wideband or high dynamic range. Different radar functions put a greater emphasis on one or the other of these parameters. For example, imaging radars put a premium on wide bandwidth, whereas pulse doppler radars require high dynamic range. Because radars are often required to operate in a variety of modes with differing bandwidth and dynamic range requirements, it is not uncommon to use different types of A/D converter, sampling at different rates for these different modes.

Data Formats. The most frequently used digital formats for A/D converters are 2's complement and offset binary.[10]

The 2's complement is the most popular method of digital representation of signed integers and is calculated by complementing every bit of a given number and adding one.

The most significant bit is referred to as the sign bit. If the sign bit is 0, the value is positive; if it is 1, the value is negative. The representation of voltage in 2's complement form is given by

$$E = k(-b_N 2^{N-1} + b_{N-1}2^{N-2} + b_{N-2}2^{N-3} + \cdots + b_1 2^0) \tag{3.36}$$

where E = analog voltage
N = number of binary digits
b_i = state of ith binary digit
k = quantization voltage

Offset binary is an alternate coding scheme in which the most negative value is represented by all zeros and the most positive value is represented by all ones. Zero is represented by a most significant bit (MSB) of one followed by all zeros. The representation of voltage in offset binary is given by

$$E = k[(b_{N-1})2^{N-1} + b_{N-1}2^{N-2} + b_{N-2}2^{N-3} + \cdots + b_1 2^0] \tag{3.35}$$

The Gray code[10] is also used in certain high-speed A/D converters in order to reduce the impact of digital output transitions on the performance of the A/D converter. The Gray code allows all adjacent transitions to be accomplished by the change of a single digit only.

Delta-Sigma Converters. Delta-sigma converters differ from conventional Nyquist rate converters by combining oversampling with noise-shaping techniques to achieve improved SNR in the bandwidth of interest. Noise shaping may be either lowpass or bandpass depending on the application. Delta-sigma architectures provide potential improvements in spurious-free dynamic range (SFDR) and SNR over conventional Nyquist converters where tight tolerances are required to achieve very low spurious performance. Digital filtering and decimation is required to produce data rates that can be handled by conventional processors. This function is either performed as an integral part of the A/D converter function or can be integrated into the digital downconversion function used to generate digital I and Q data, as described in Section 6.11.

Performance Characteristics. The primary performance characteristics of A/D converters are the sample rate or usable bandwidth and resolution, the range over which the signals can be accurately digitized. The resolution is limited by both noise and distortion and can be described by a variety of parameters.

Sample Rate. Sampling of band-limited signals is performed without aliasing distortion, provided that the sample rate (f_s) is greater than twice the signal bandwidth and provided the signal bandwidth does not straddle the Nyquist frequency ($f_s/2$) or any integer multiple ($Nf_s/2$).

In conventional baseband approaches, sampling is usually performed at the minimum rate to meet the Nyquist criteria. Since the baseband I and Q signals have bandwidths ($B/2$) equal to half the IF signal bandwidth, a sample rate just greater than the IF bandwidth is required (see Figure 6.14).

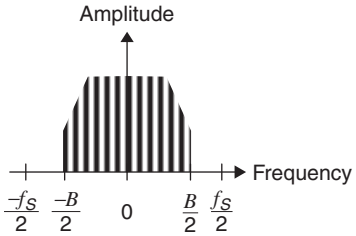


FIGURE 6.14 Baseband sampling

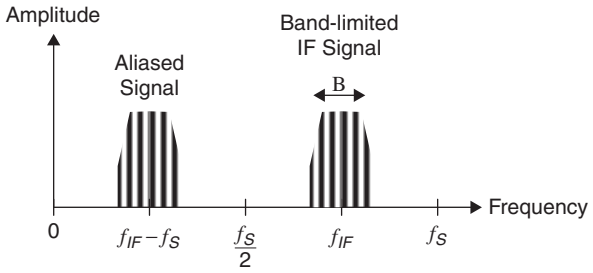


FIGURE 6.15 IF sampling in second Nyquist region

For IF sampling, a frequency at least twice the IF bandwidth is required; however, oversampling is typically employed to ease alias rejection filtering and to reduce the effect of A/D converter quantization noise. IF sampling is often performed with the signal located in the second Nyquist region, as shown in Figure 6.15 or in higher Nyquist regions.

Stated Resolution. The stated resolution of an A/D converter is the number of output data bits per sample. The full-scale voltage range of a Nyquist rate converter is given by $V_{FS} = 2^N Q$, where N is the stated resolution and Q is the least significant bit (LSB) size.

Signal-to-Noise-Ratio (SNR). SNR is the ratio of rms signal amplitude to rms A/D converter noise power. For an ideal A/D converter, the only error is due to quantization. Provided that the input signal is sufficiently large relative to the quantization size and uncorrelated to the sampling signal, the quantization error is essentially random and is assumed to be white. The rms quantization noise is $Q/\sqrt{12}$, and signal-to-quantization-noise ratio (SQNR) of an ideal A/D converter is given by

$$SQNR(dB) = 6.02N + 1.76 \tag{6.37}$$

Practical A/D converters have additional sampling errors other than quantization, including thermal noise and aperture jitter. Provided that these additional errors can be characterized as white, they can be combined with the quantization noise with a resulting SNR less than the theoretical SNR of the ideal converter. Because various A/D converter error mechanisms are dependent on input signal level and frequency, it is important to characterize devices over the full range of input conditions to be expected. The available signal-to-noise ratio of state-of-the-art high-speed A/D converters has been shown[11] to fall off by one-bit (6 dB) for every doubling of the sample rate. Over-sampling of the signal followed by filtering and decimation provides an improvement of one half-bit (3 dB) in the achievable signal-to-noise-ratio for each doubling of the sample rate. Thus, for high dynamic-range applications, the best performance is achieved using a state-of-the-art A/D converter that has a maximum sample rate just sufficient for the application.

Spurious Free Dynamic Range (SFDR). SFDR is the ratio of the single-tone signal amplitude to the largest spurious signal amplitude and is usually stated in dB. Similar to SNR, the spurious performance of an A/D converter is dependent on the

input signal frequency and amplitude. The frequency of spurious signals is also dependent on the input signal frequency with the highest values typically due to low order harmonics or their aliases. When using IF sampling with a significant over-sampling ratio ($f_s \gg B/2$), the worst spurious signals may be avoided by choosing the sample frequency relative to signal frequency such that the unwanted spurious signals fall outside the signal bandwidth of interest. If the worst case spurious can be avoided, the specified SFDR is less important than the levels of the specific spurious components that fall within the bandwidth of interest. Again, it is important to characterize devices over the range of expected operating conditions.

The impact of A/D converter spurious signals on radar performance depends on the type of waveforms being processed and the digital signal processing being performed. In applications using chirp waveforms with large time-bandwidth products, spurious signals are less critical as they are effectively rejected in the pulse compression process because their coding does not match that of the wanted signal. In pulse doppler applications, spurious signals are of much greater concern because they can create components with doppler at a variety of frequencies that may not be rejected by the clutter filtering.

Signal-to-Noise-and-Distortion Ratio (SINAD). SINAD is the rms signal amplitude to the rms value of the A/D converter noise plus distortion. The noise plus distortion includes all spectral components, excluding DC and the fundamental up to the Nyquist frequency. SINAD is a useful figure of merit for A/D converters, but in digital receiver applications, where the worst spurious components may fall outside of the bandwidth of interest, it is not necessarily a key discriminator between competing converters for a specific application.

Effective Number of Bits (ENOB). The term *effective number of bits* is often used to state the true performance of an A/D converter and has been stated in the literature[11] in terms of SINAD and SNR, as given below. Consequently, it is important to differentiate between definitions when using this term.

$$N_{\text{eff}} = [SINAD(dB) - 1.76] / 6.02 \tag{6.38}$$

$$N_{\text{eff}} = [SNR(dB) - 1.76] / 6.02 \tag{6.39}$$

Two Tone Intermodulation Distortion (IMD). Two tone intermodulation distortion is also important in receiver applications. Testing is performed with two sinusoidal input signals of unequal frequency and levels set such that the sum of the two inputs does not exceed the A/D converter full-scale level. Similar to IMD for amplifiers, the most significant distortion is usually second order or third order IMD products. However, due to the complex nature of the distortion mechanism in A/D converters, the amplitude of IMD products is not easily characterized and predicted by the measurement of an input intercept point.

Input Noise Level and Dynamic Range. Accurate setting of the A/D converter input noise level relative to the A/D converter noise is critical to achieving the optimum trade-off between dynamic range and system noise floor. Too high a level of noise into the A/D converter will degrade the available dynamic range; too low a level will degrade the overall system noise floor. Sufficient total noise should be applied to the A/D converter input to randomize or "whiten" the quantization noise.

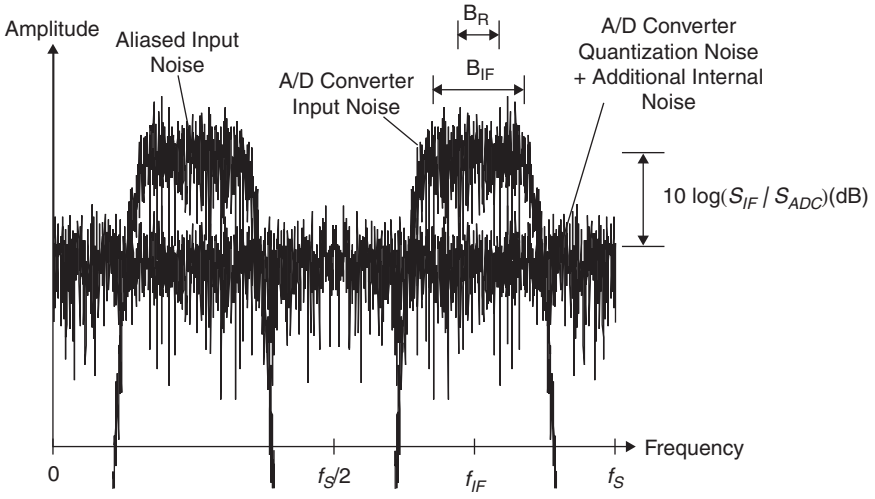


FIGURE 6.16 IF sampling noise spectrums

This can be achieved with rms input noise (σ) equal to the LSB step size (Q). In addition, the input noise power spectral density should be sufficient to minimize the impact on system noise due to the A/D converter noise. The impact on overall noise due to quantization noise is given by[7]

$$\frac{\tilde{\sigma}^2}{\sigma^2} = 1 + \frac{Q^2}{12\sigma^2} \qquad \sigma \geq Q \tag{6.40}$$

Typical operating points are in the range of $\sigma/Q = 2$ to $\sigma/Q = 1$, with corresponding noise power degradation due to quantization of 0.09 dB and 0.35 dB, respectively.

In practice, the SNR of high-speed converters is often such that the noise of the A/D converter is significantly greater than the theoretical quantization noise. In addition, the A/D converter input signal noise bandwidth may be significantly less than the Nyquist bandwidth. This is a significant factor in IF sampling applications where the IF noise bandwidth is often less than ¼ of the Nyquist bandwidth. In this case, the total input and A/D converter noise must be sufficient to whiten the quantization noise, and the power spectral density of the input noise should be sufficiently greater than that of the A/D converter, as illustrated in Figure 6.16. In some cases, out-of-band noise may be added to whiten the A/D converter quantization noise and spurious signals. The out-of-band noise is then rejected through subsequent digital signal processing.

The resulting SNR of the system after digital filtering with receiver bandwidth B_R and sample rate f_s is given by

$$SNR_{\text{SYS}}(dB) = SNR_{\text{ADC}}(dB) + 10\log_{10}\left(\frac{f_S}{2B_R}\right) - 10\log_{10}\left(1 + S_{\text{IF}}/S_{\text{ADC}}\right) \tag{6.41}$$

where $S_{\text{IF}}/S_{\text{ADC}}$ is the ratio of noise power spectral density of the A/D converter input signal to the power spectral density of the A/D converter. The degradation of overall sensitivity due to the A/D converter noise is given by

$$L(dB) = 10\log_{10}(1 + S_{\text{ADC}}/S_{\text{IF}}) \tag{6.42}$$

A/D Converter Sample Clock Stability. The stability of the sample clock is critical to achieving the full capability of an A/D converter. Sample-to-sample variation in the sampling interval, called *aperture uncertainty* or *aperture jitter*, produces a sampling error, proportional to the rate of change of input voltage. For a sinusoidal input signal, the SNR due to aperture uncertainty alone is given by[12]

$$SNR(dB) = -20\log_{10}(2\pi f\sigma_j) \tag{6.43}$$

where f = input signal frequency
σ_j = rms aperture jitter

Similarly, close-to-carrier noise sidebands present on the sample clock signal are transferred to sidebands on the sampled input signal, reduced by $20\log_{10}(f/f_S)$ dB. For example, in an IF sampling application with the input signal ¾ of the sample frequency, the close-to-carrier phase noise of the sample clock will be transferred to the output of the A/D converter output data signal, reduced by 2.5 dB.

6.11 DIGITAL RECEIVERS

The availability of high-speed analog-to-digital converters capable of direct sampling of radar receiver IF signals has resulted in the almost universal adoption of digital receiver architectures over conventional analog I/Q demodulation. In a digital receiver, a single A/D converter is used to digitize the received signal, and digital signal processing is used to perform the downconversion to I and Q baseband signals. Continuing advances in sampling speeds are leading to sampling at increasing frequencies, sometimes eliminating the need for a second downconversion, with the possibility approaching of sampling directly at the radar RF frequency. The benefits of IF sampling over conventional analog I/Q demodulation are

- Virtual elimination of I and Q imbalance
- Virtual elimination of DC offset errors
- Reduced channel-to-channel variation
- Improved linearity
- Flexibility of bandwidth and sample rate
- Tight filter tolerance, phase linearity, and improved anti-alias filtering
- Reduced component cost, size, weight, and power dissipation

The use of a high IF frequency is desirable as it eases the downconversion and filtering process; however, the use of higher frequencies places greater demands on the performance of the A/D converter. Direct RF sampling is considered the ultimate goal of digital receivers, with all the tuning and filtering performed through digital signal processing. The advantage being the almost complete elimination of analog hardware. However, not only does the A/D converter have to sample the RF directly, but unless it is preceded by tunable RF preselector filters, the A/D converter input must have the dynamic range to handle all of the signals present in the radar band simultaneously. Generally, the interference power entering the A/D converter is proportional to the bandwidth of components in front of the

A/D converter. The required A/D converter SNR to avoid saturation on the interfering signals is given by

$$SNR_{\text{ADC}}(dB) = 10\log_{10}\left(\frac{P_I C^2}{N_{\text{ADC}}}\right) \tag{6.44}$$

where

P_I = interference power at A/D converter input
C = crest factor
N_{ADC} = A/D converter noise

The crest factor is the peak level that can be handled within the full-scale range of the A/D converter relative to the rms interference level. It is set to achieve a sufficiently high probability that full-scale will not be exceeded. For example, with gaussian noise, a crest factor of 4 sets the peak level at the 4σ level (12 dB above the rms level) with a probability of 0.999937 that the full-scale is not exceeded on each A/D converter sample.

Setting the system noise level power spectral density into the A/D converter $R(dB)$ above the A/D converter noise gives

$$R(dB) = 10\log_{10}\left(\frac{f_s N_{\text{SYS}}}{2B_{\text{IF}} N_{\text{ADC}}}\right) \tag{6.45}$$

where

N_{SYS} = system noise at A/D converter input in bandwidth B_{IF}

Combining Eq. 6.43 and 6.44 gives the required SNR as

$$SNR_{\text{ADC}}(dB) = 10\log_{10}\left(\frac{2P_I C^2 B_{\text{IF}}}{f_S N_{\text{SYS}}}\right) + R(dB) \tag{6.46}$$

The generation of baseband I and Q signals from the IF sampled A/D converter data is performed using digital signal processing and can be implemented through a variety of approaches.[7] Two approaches are described next.

Digital Downconversion. The digital downconversion approach is shown in Figure 6.17. The signal is sampled by the A/D converter, frequency shifted to baseband, low-pass filtered, and decimated to produce I/Q digital data. The signal spectrum at each stage of the process is shown in Figure 6.18. In continuous-time (Fig. 6.18*a*), frequency is in hertz and is represented by F. In discrete-time (Fig. 6.18*b–e*), frequency is in radians per sample and is represented by ω. The spectrum of the analog input signal $x(t)$ is shown in Figure 6.18*a*, with the signal spectrum centered at F_0 hertz. The signal is sampled by the A/D converter at frequency F_s, producing the time sequence $\tilde{x}(n)$ and frequency spectrum $\tilde{X}(\omega)$ centered at frequency ω_0 with the image centered at $-\omega_0$. The A/D converter output signal is then frequency shifted by complex multiplication with the reference signal $e^{-j\omega_0 n}$, corresponding to a reference signal rotating at ω_0 radians per sample, centering the signal spectrum $\bar{X}(\omega)$ about zero. The unwanted image is re-centered at $-2\omega_0$ if $\omega_0 > \pi/2$ or $-2\omega_0 + 2\pi$ if $\omega_0 \le \pi/2$. The unwanted image is then rejected using the FIR filter with impulse response $h(n)$ producing output $\hat{x}(n)$ with spectrum $\hat{X}(\omega)$. Finally, the sample rate is reduced by

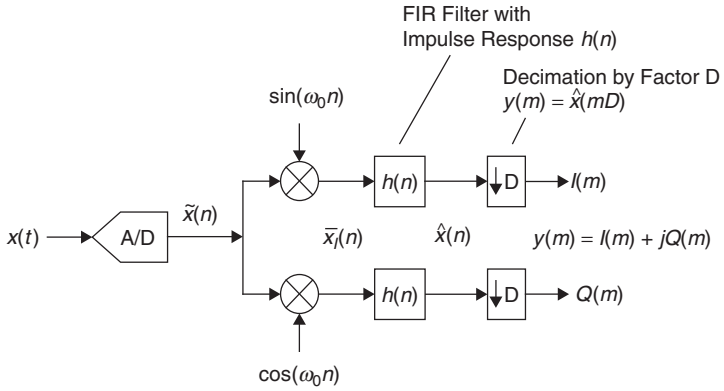


FIGURE 6.17 Digital downconversion architecture

selecting every Dth sample. Provided the filter response $H(\omega)$ has sufficient rejection for frequencies $|\omega| \geq \pi/D$, there will be negligible aliasing and loss of information in the decimation process.

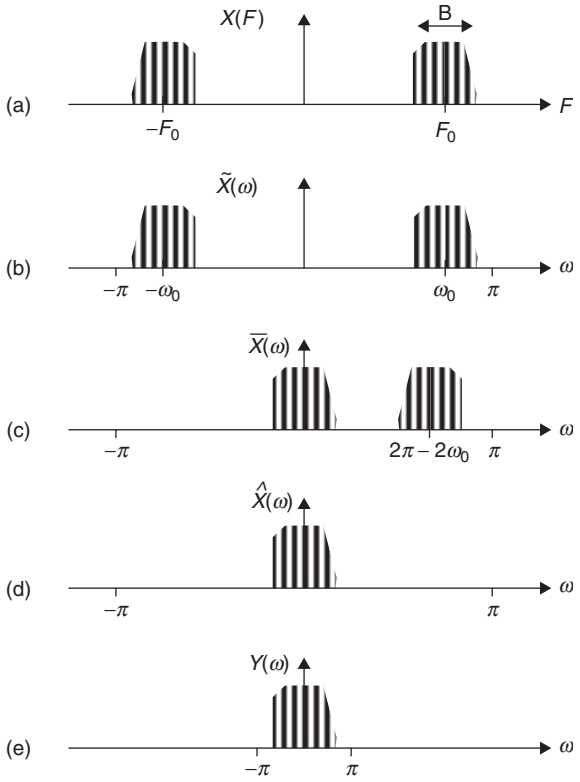


FIGURE 6.18 Digital downconversion spectra

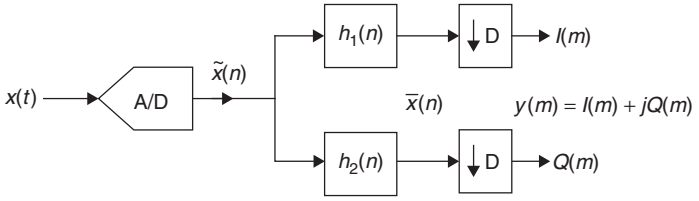


FIGURE 6.19 Hilbert transformer architecture

Hilbert Transformer. An alternative digital receiver architecture is shown in Figure 6.19 with the relevant signal spectra shown in Figure 6.20. The A/D converter output signal $\tilde{x}(n)$ is processed using a Hilbert transformer comprising FIR filters $h_1(n)$ and $h_2(n)$, where the frequency responses are given by

$$|H_1(\omega)| \approx |H_2(\omega)| \approx 1 \qquad |\omega - \omega_0| \le B \tag{6.47}$$

and

$$\frac{H_1(\omega)}{H_2(\omega)} \approx \begin{cases} -j, & |\omega - \omega_0| \le B \\ j, & |\omega + \omega_0| \le B \end{cases} \tag{6.48}$$

The filter outputs form the desired complex valued signal $\bar{x}(n)$ centered at frequency ω_0, while rejecting the image centered at $-\omega_0$. The final stage is to perform a frequency shift and sample rate reduction by decimating the signal by selecting every Dth sample.

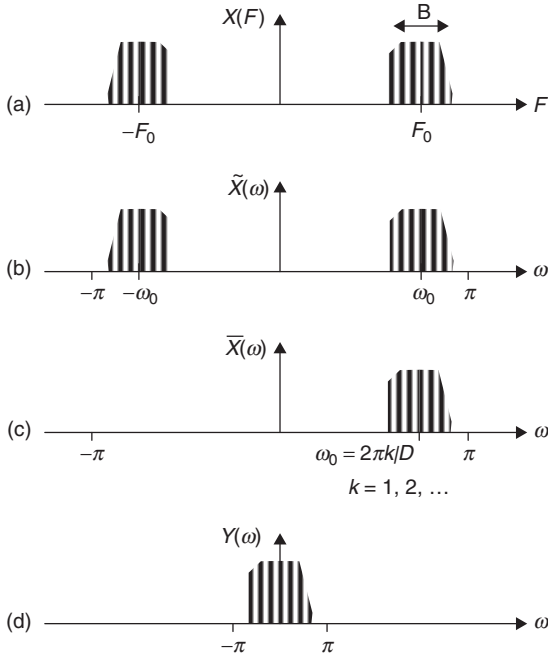


FIGURE 6.20 Spectra of Hilbert transformer receiver

If the spectrum of $\bar{X}(\omega)$ is centered at frequency $\omega_0 = 2\pi k/D$, $k = 1,2,..,$ the decimation will center the spectrum $Y(\omega)$ about zero. Provided the filter responses have sufficient rejection for frequencies $|\omega \pm \omega_0| \geq \pi/D$, there will be negligible aliasing and loss of information in the decimation process.

I/Q Errors. Digital *I* and *Q* generation does not produce signals without error, as is often stated, but instead allows the generation of these signals with errors that are sufficiently small to be considered negligible. The primary cause of the imbalance is the non-ideal filter responses. An infinite number of taps would be required to set the passband gain to unity and the stopband gain to zero; however, for most applications, sufficient processing resources are available to reduce the errors to insignificant levels. Finite length words for filter coefficients produce non-ideal filter responses. The effect on passband response is typically negligible, but significant distortion of the filter stopband rejection can occur, potentially effecting I/Q balance.

Digital Downconversion Using Multirate Processing and Polyphase filters. There are many variations to these basic approaches, and specific implementations often utilize efficient approaches that minimize the number of calculations required with emphasis on reducing the number of multiplications, as these require significantly more resources than additions. Two techniques used to reduce the FIR filter processing burden are multirate processing and polyphase filtering.[13] The digital downconversion approach is shown in Figure 6.21 using multirate processing. The first FIR filter $h_1(n)$ provides sufficient reduction to prevent aliasing in the first decimation by factor D_1, the second filter $h_2(n)$ provides alias reduction for the second decimation and can also be used to correct passband ripple or droop due to filter $h_1(n)$. For large decimation factors, more than two decimation stages may be used.

A popular filter for the first stage is the Cascaded Integrator Comb[14] (CIC) decimator filter that can be implemented without multipliers. These filters provide rejection in the stopband at frequencies that alias to the passband as a result of decimation. Since they provide relatively large passband droop and slow stopband rejection, they are generally followed by a FIR filter that can both correct for CIC passband droop and

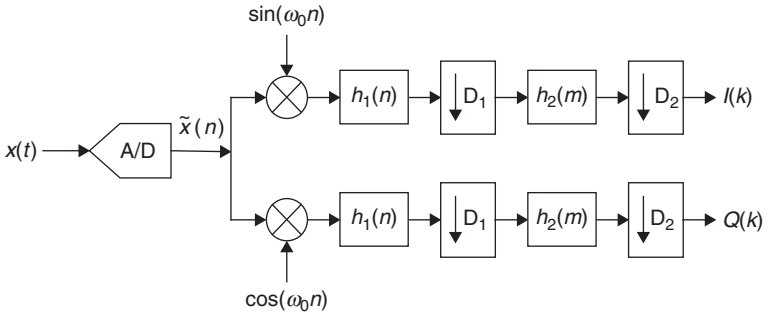


FIGURE 6.21 Digital downconversion architecture

provide the desired stopband rejection response. The *k*th order CIC filter for decimation factor *D* has transfer function:

$$H_K(z) = \left[\sum_{m=0}^{D-1} z^{-m}\right]^K = \left[\frac{1-z^{-D}}{1-z^{-1}}\right]^K \tag{6.49}$$

A polyphase filter is a filter bank that splits an input signal into *D* sub-band filters operating at a sample rate reduced by a factor *D*, providing a computationally efficient approach to performing the FIR filtering followed by decimation in a digital receiver. Rather than computing all the filter output samples and only using every *D*th sample, the polyphase approach calculates only those that are actually used. Figure 6.22 and Eq. 6.50 define how the filter with impulse response $h(n)$, followed with decimation by factor *D*, is implemented in a polyphase structure. The input signal $x(n)$ is divided into *D* parallel paths by the "commutator," which outputs samples in turn, rotating in a counterclockwise direction, to each of the FIR filters operating at the reduced sample rate. The outputs of the FIR filters are summed to produce the output signal $y(m)$. This architecture is beneficial as it provides an approach that can be easily parallelized at rate F_X/D.

$$p_k(n) = h(k + nD) \qquad \begin{matrix} \mathrm{k} = 0, 1, \ldots, D-1 \\ n = 0, 1, \ldots, K-1 \end{matrix} \tag{6.50}$$

Multi-Channel Receiver Considerations. Modern radar systems rarely contain only one receiver channel. Monopulse processing, for example, requires two or more channels to process sum and delta signals. Additionally, the channels must be coherent, synchronized in time, and well matched in phase and amplitude. Digital beamforming systems require a large number of channels with similar coherence and synchronization requirements and tight phase and amplitude tracking. The coherence requirement dictates the relative phase stability of LO and A/D converter clock signals used for each receive channel. The time synchronization requirement means that A/D converter clock signals for each channel must be aligned in time and decimation must be performed in phase for each channel. Phase and amplitude imbalance between channels is a result of variation in the

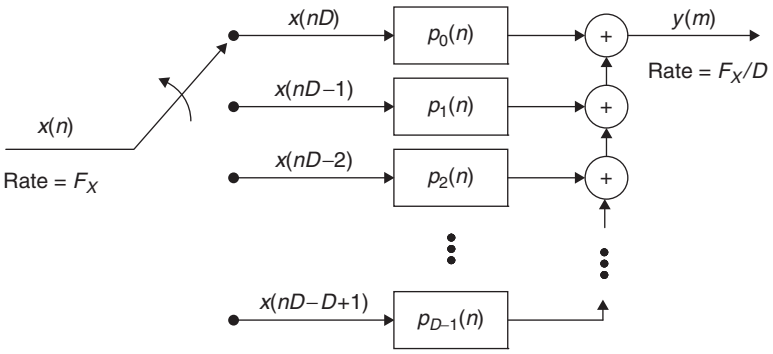


FIGURE 6.22 Decimation using polyphase filters

analog circuitry prior to and within the A/D converter. If the IF filter bandwidth is wide relative to the digital receiver bandwidth, the majority of the error between channels will be a constant gain and phase offset across the receiver bandwidth. A single correction, applied as a complex multiplication of *I*/*Q* data, will correct for gain and phase offsets and is usually adequate to provide the required channel tracking for monopulse applications. When tighter channel tracking is required, such as for sidelobe canceler or digital beamforming applications, FIR filter equalization can be used to correct for frequency dependent variations across the receiver bandwidth. FIR filter equalization can be performed either subsequent to the FIR filtering used to generate *I*/*Q* data or combined with these filters. It should be noted that to correct for frequency and phase variation across the receiver bandwidth requires FIR filters with complex coefficients, applied equally to *I* and *Q* data. Real value coefficients typically used in I/Q generation provide filter responses symmetrical about zero frequency. Correction of IF filter frequency response errors will, in general, require asymmetric frequency correction that can only be provided at baseband using complex coefficients.

The degree to which these multiple receiver channels must track depends on the specific system requirements. Although modern systems typically include some degree of channel equalization function, a reasonable degree of tracking between gain, phase, and timing must be maintained in order to allow the channel equalization to be performed using digital signal processing without consuming excessive processing resources. Also, the relative stability of the radar channels as a function of time and temperature must be such that the corrections can maintain adequate tracking during the time between calibration intervals.

Digital beamforming systems require a large number of receiver channels. In these applications, size, weight, power dissipation, and cost are critical considerations.

6.12 DIPLEX OPERATION

Diplex Benefits. Diplex operation consists of two receivers that simultaneously process returns from transmissions on different frequencies. Transmissions are usually non-overlapping in time to avoid a 6 dB increase in peak power and because most radar transmitters are operated in saturation and simultaneous transmission at multiple frequencies would produce significant transmitted intermodulation distortion.

The sensitivity benefit of diplex operation for detecting Swerling 1 targets is shown in Figure 6.23, increasing with probability of detection (P_D). For example, diplex operation achieves 90% P_D with 2.6 dB less total signal power than simplex. Assumptions made in deriving Figure 6.23 are

1. Returns on the two frequencies are added in voltage or power prior to the detection decision rather than being subjected to individual detection decisions.
2. Separation of the two frequencies is sufficient to make their Swerling 1 fluctuations independent. This depends on the physical length of the target in the range dimension λ_R. The minimum frequency separation is 150 MHz/λ_R (m); 25 MHz will maintain the diplex benefit for aircraft longer than 6 m (20 ft).
3. Equal energy is transmitted in both pulses. A 2:1 imbalance sacrifices only 0.2 dB of the benefit at 90% P_D.

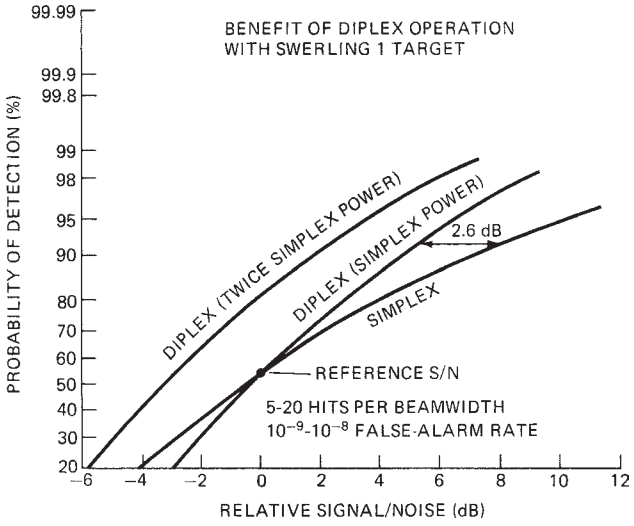


FIGURE 6.23 Diplex operation improves the sensitivity of the receiver

Both linear and asymmetrical nonlinear FM produce a range error as a function of doppler due to range-doppler coupling. These range displacements must match in the two receivers to within a small fraction of the compressed pulse width; otherwise, the sensitivity benefits of diplex operation are not fully achieved and range accuracy may be degraded.

Implementation. Diplex operation can be implemented with a variety of approaches. Complete replication of the receiver channels is typically the most expensive approach and may be required if the frequency separation is very large. A more common approach is separation of the frequencies at the first IF, as this does not require complete duplication of the RF front end or the first LO signal. Separate second local oscillator or I/Q demodulator reference frequencies can be used to process the different frequencies. With the use of high-speed IF sampling, it is also possible to digitize both signals simultaneously using a single A/D converter and perform the frequency separation using digital signal processing. Whichever approach is used, care must be taken to provide adequate dynamic range and linearity to prevent intermodulation distortion from degrading radar performance.

6.13 WAVEFORM GENERATION AND UPCONVERSION

The exciter function of waveform generation and upconversion is often tightly coupled with the receiver function. The requirement for coherence between the receiver and exciter is a major factor for this tight coupling and the use of the same LO frequencies within the receiver and exciter usually results in hardware savings. Similar to the migration to digital receiver architectures, the exciter functionality is increasingly being implemented using digital approaches.

Direct Digital Synthesizer. The Direct Digital Synthesizer[15] (DDS) produces waveforms using digital techniques and provides significant improvements in stability, precision, agility, and versatility over analog techniques. The main limitations are the noise and spurious signals as described below. The general DDS architecture is shown in Figure 6.24. The double accumulator architecture, comprising the frequency and phase accumulators, enables the generation of CW, linear FM (chirp), nonlinear (piece-wise linear) FM, frequency modulated, and phase modulated waveforms. CW waveforms are generated by applying a constant frequency word (digitized frequency representation) input to the phase accumulator, creating a linear phase sequence that is first truncated then input to a cosine (or sine) lookup table that outputs the corresponding sinusoidal signal value to the digital-to-analog (D/A) converter. The frequency resolution is dependent on the number of bits and the clock frequency of the phase accumulator. The output frequency is given by

$$f_{\text{out}} = \frac{M_f f_{\text{clk}}}{2^{N_\phi}} \tag{6.51}$$

where

M_f = frequency word, input to the phase accumulator
f_{clk} = phase accumulator clock frequency
N_ϕ = number of bits of phase accumulator

Linear FM or chirp waveforms are generated by applying a constant chirp slope word (digitized chirp slope representation) to the input of the frequency accumulator, creating a quadratic phase sequence at the output of the phase register. Piecewise-linear or nonlinear FM waveforms can be generated by applying a time-varying slope input to the frequency register. The frequency accumulator may be clocked either at the same rate as the phase accumulator or at a sub-multiple to provide finer chirp slope resolution. If both accumulators are clocked at the same rate, the chirp slope is given by

$$\frac{\Delta f_{\text{out}}}{\Delta t} = \frac{M_S f_{\text{clk}}^2}{2^{N_f}} \tag{6.52}$$

where

M_S = chirp slope word, input to the frequency accumulator
N_f = number of bits of frequency accumulator

Frequency modulated and phase modulated waveforms can be created applying time-varying inputs to the frequency modulation (FM) and phase modulation (PM) ports.

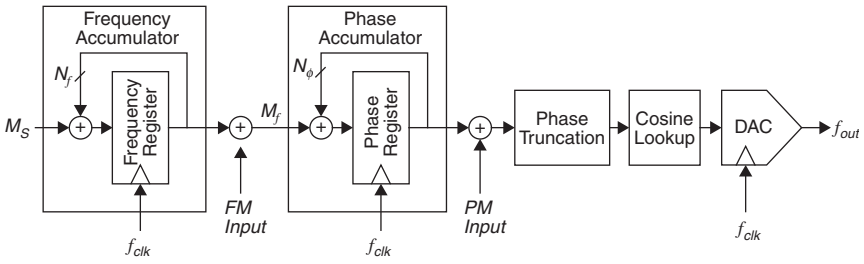


FIGURE 6.24 Direct Digital Synthesizer block diagram

Errors such as phase truncation and D/A converter quantization and nonlinearity produce spurious signals due to their deterministic nature. The spurious signal frequencies generated by a DDS can be readily predicted[16] as they are a function of the digital architecture and programmed frequency. The spurious signal magnitudes are less predictable as the magnitudes of the dominant spurious signals are a function of the D/A converter nonlinearity.

When generating CW waveforms, the D/A converter sequence repeats after 2^K samples where 2^K equals the greatest common divisor of $2^{N\phi}$ and M_f. Thus, spurious signals occur only at frequencies:

$$f_{\text{spur}} = \frac{nf_{\text{clk}}}{2^K}, \quad \text{n} = 0,1,2,... \tag{6.53}$$

In the extreme case where M_f does not contain the factor 2, this creates a spurious frequency spacing of $f_{\text{clk}}/2^{N\phi}$. For example, with a 1 GHz clock and 32-bit frequency accumulator, the spurious frequency spacing can be as close as 0.23 Hz. In most cases, such closely spaced spurious signals cannot be differentiated from noise. Conversely, choosing values of M_f that contain large factors of 2^N creates relatively large spurious spacing. For example, using a 640 MHz clock allows the generation of frequencies at multiples of 10 MHz with all the spurious components occurring at multiples of 10 MHz.

The impact of DDS spurious signals on radar performance depends on the nature of the spurious signals and the type of radar processing involved. Applications using chirp waveforms with large time-bandwidth products are typically less sensitive to DDS spurious signals since the DDS spurious signals chirp at a different rate to that of the wanted signal. The spurious signals are thus rejected during pulse compression. In pulse doppler applications, spurious signals are of much greater concern; however, their effects can be mitigated by ensuring that the DDS generates each waveform from the same initial conditions. Restarting the DDS for every pulse guarantees that the same digital sequence will be input to the D/A converter for each pulse. The result is a DDS output that only contains spectral components at multiples of the PRF.

Techniques have been proposed or incorporated into DDS devices that reduce spurious levels by adding dithering to reduce the effects of limited word lengths. The effect of these techniques and the spurious signals that they are designed to mitigate should be considered carefully as they may be detrimental to radar performance. The use of dithering will randomize the spurious signal, resulting in pulse-to-pulse variations in the digital sequence output to the D/A converter, a result that is undesirable in pulse doppler applications.

Truly random errors are not generated by the digital portion of the DDS. The only nondeterministic errors are a result of the D/A converter performance in the form of internal clock jitter or additive thermal noise and the effect of the phase noise on the input clock signal.

Internal D/A converter clock jitter produces phase modulation of the output signal proportional to the output frequency. Similarly, phase noise present on the clock input signal is transferred to the output signal, reduced by $20\log_{10}(f_{\text{out}}/f_{\text{clk}})$ dB. D/A converter additive thermal noise is independent of output signal frequency and produces both phase and amplitude noise components.

Frequency Multipliers. Frequency multiplication allows signals to be increased in both frequency and bandwidth. Frequency multiplication is frequently used in generating local oscillator CW frequencies where all frequencies are typically based on a

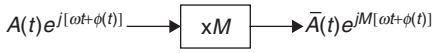


FIGURE 6.25 Frequency multiplier operation

low frequency reference. They also provide the capability for wide-bandwidth chirp waveforms that cannot be generated directly using available DDS devices. Frequency multipliers operate as shown in Figure 6.25, by multiplying the phase of the input signal by the integer multiplication factor M. Since in practice the process typically includes some form of limiting, the output amplitude $\bar{A}(t)$ generally has a lower amplitude variation than the input signal amplitude $A(t)$.

Because the multiplication process multiplies up the variations in the signal phase by factor M, input phase noise and spurious phase modulations are increased by $20\log_{10}(M)$ dB. Similarly, variations in the phase of the signal as a function of frequency are multiplied up. These variations are produced during signal filtering and may be present on the input signal. For chirp waveforms, this can result in a significant degradation in the range sidelobe performance. Also, practical multipliers may have a significant phase variation as a function of frequency. If the input signal phase distortion is given by

$$\phi(f) = \beta \sin\left(\frac{2\pi nf}{B}\right) \tag{6.54}$$

where

β = peak phase ripple
B = waveform input bandwidth
n = number of cycles of phase ripple

the resulting output distortion produces range sidelobes at times $\pm n/MB$ and magnitude $20\log_{10}(M\beta/2)$ relative to the main beam of the target return. As an example, generating a chirp waveform that has range sidelobes better than 35 dB using an ×8 multiplier requires that the input signal has less than 0.5 degrees peak-peak phase ripple.

Frequency multipliers can be implemented using a variety of techniques, such as using step recovery diode multipliers or using phase locked loops. Where wide percentage bandwidth and fast settling is required, the most common technique is to cascade a series of frequency doublers or low order multipliers. This type of multiplier can also provide near ideal phase noise performance, but has significant phase modulation as a function of frequency as it contains filters between each stage of multiplication.

Predistortion of the multiplier input waveform is often used in order to produce wideband chirp waveforms with low range sidelobe performance. If the multiplier is characterized by an output phase distortion as a function of input frequency given by $\phi(\omega)$, then a predistortion of the input signal by phase $-\phi(\omega)/M$ will equalize the multiplier response. Predistortion can be performed very precisely by adding the phase modulation via the DDS that is used to generate the chirp waveform.

Waveform Upconversion. Upconversion of exciter waveforms is similar to downconversion within the receiver. Also, similar practical considerations of mixer spurious and image rejection apply. The one significant additional challenge is the rejection of the LO leakage. LO rejection typically imposes tight filter rejection requirements on the RF filters, and for wide-tunable ranges, switched filters are often required.

REFERENCES

1. M. I. Skolnik, *Radar Handbook*, 2nd Ed., New York: McGraw Hill, 1990.
2. M. I. Skolnik, *Radar Handbook*, 1st Ed., New York: McGraw Hill, 1971.
3. R. E. Watson, "Receiver dynamic range: Part 1," Watkins Johnson Company, Technical Note, vol. 14, no 1, January/February 1987.
4. B. C. Henderson, "Mixers in microwave systems (Part 1)," Watkins Johnson Company, Technical Note, vol.17, no.1, January/February 1990.
5. D.W. Allan, H. Hellwig, P. Kartaschoff, J. Vanier, J. Vig, G. M. R. Winkler, and N. F. Yannoni, "Standard terminology for fundamental frequency and time metrology," in *Proceedings of the 42nd Annual Frequency Control Symposium*, Baltimore, MD, June 1–4, 1988, pp. 419–425.
6. P. Renoult, E. Girardet, and L. Bidart, "Mechanical and acoustic effects in low phase noise piezo-electric oscillators," presented at IEEE, 43rd Annual Symposium on Frequency Control, 1989.
7. M. A. Richards, *Fundamental of Radar Signal Processing*, New York: McGraw-Hill, 2005.
8. A. I. Zverev, *Handbook of Filter Synthesis*, New York: John Wiley and Sons, Inc., 1967.
9. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, New York: Prentice Hall Inc., 1989.
10. W. Kester, *The Data Conversion Handbook*, London: Elsevier/Newnes, 2005.
11. R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, April 1999.
12. B. Brannon, "Sampled systems and the effects of clock phase noise and jitter," Analog Devices Inc., Application Note, AN-756, 2004.
13. J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 2nd Ed., New York: Macmillan, 1992.
14. E. B. Hogenauer, "An economical class of digital Filters for decimation and interpolation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-29, no. 2, April 1981.
15. J. Tierney, C. M. Radar, and B. Gold, "A digital frequency synthesizer," *IEEE Trans*. AU-19, pp.43–48, March 1971.
16. H. T. Nicholas III and H. Samueli, "An analysis of the output spectrum of direct digital frequency synthesizers in the presence of phase-accumulator truncation," *Proceedings. 41st annual Frequency Control Symposium*, USERACOM, Ft. Monmouth, NJ, May 1987, pp. 495–502.

Chapter 7

Automatic Detection, Tracking, and Sensor Integration

W. G. Bath and G. V. Trunk

The Johns Hopkins University Applied Physics Laboratory

7.1 INTRODUCTION

As digital processing has increased in speed and digital hardware has decreased in cost and size, radars have become more and more automated, so that automatic detection and tracking (ADT) systems are associated with almost all but the simplest of radars.

In this chapter, automatic detection, automatic tracking, and sensor integration techniques for surveillance radars are discussed. Included in the discussion are various noncoherent integrators that provide target enhancement, thresholding techniques for false alarms and target suppression, and algorithms for estimating target position and resolving targets. Then, an overview of the entire tracking system is given, followed by a discussion of its various components such as track initiation, correlation logic, tracking filter, and maneuver-following logic. Finally, the chapter concludes with a discussion of sensor integration and radar netting, including both colocated and multisite systems.

7.2 AUTOMATIC DETECTION

In the 1940s, Marcum[1] applied statistical decision theory to radar and later Swerling[2] extended the work to fluctuating targets. They investigated many of the statistical problems associated with the noncoherent detection of targets in gaussian noise. (Note: If the inphase and quadrature components are gaussian distributed, the envelope is Rayleigh distributed and the power is exponentially distributed.) Marcum's most important result was the generation of curves of probability of detection (P_D) versus signal-to-noise ratio (S/N) for a detector that sums N envelope-detected samples (either linear or square-law) under the assumption of equal signal amplitudes. Whereas for a phased array, the equal amplitude assumption is valid; for a rotating radar, the returned signal amplitude is modulated by the antenna pattern as the beam sweeps over

the target. Many authors have investigated various detectors, comparing detection performance and angular estimation results with optimal values, and many of these results are presented later in this section.

In the original work on detectors, the environment was assumed known and homogeneous, so that fixed thresholds could be used. However, a realistic radar environment (e.g., containing land, sea, and rain) will cause an exorbitant number of false alarms for a fixed-threshold system that does not utilize excellent coherent processing. Three main approaches—adaptive thresholding, nonparametric detectors, and clutter maps—have been used to solve the noncoherent, false-alarm problem. Both adaptive thresholding and nonparametric detectors are based on the assumption that homogeneity exists in a small region about the range cell that is being tested. The adaptive thresholding method assumes that the noise density is known except for a few unknown parameters (e.g., the mean and the variance). The surrounding reference cells are then used to estimate the unknown parameters, and a threshold based on the estimated density is obtained. Nonparametric detectors obtain a constant false-alarm rate (CFAR) by ranking (ordering the samples from smallest to largest) the test sample with the reference cells. Under the hypothesis that all the samples (test and reference) are independent samples from an unknown density function, the rank of the test sample is uniform; and consequently, a threshold that yields CFAR can be set. Clutter maps store an average background level for each range-azimuth cell. A target is then declared in a range-azimuth cell if the new value exceeds the average background level by a specified amount.

Optimal Detector. The radar detection problem is a binary hypothesis-testing problem in which H_0 denotes the hypothesis that no target is present and H_1 is the hypothesis that the target is present. While several criteria (i.e., definitions of optimality) can be used to solve this problem, the most appropriate for radar is the Neyman-Pearson.[3] This criterion maximizes the probability of detection P_D for a given probability of false alarm P_{fa} by comparing the likelihood ratio L (defined by Eq. 7.1) to an appropriate threshold T that determines the P_{fa}. A target is declared present if

$$L(x_1,\ldots,x_n) = \frac{p(x_1,\ldots,x_n|\, H_1)}{p(x_1,\ldots,x_n|\, H_0)} \geq T \tag{7.1}$$

where $p(x_1,\ldots, x_n|H_1)$ and $p(x_1,\ldots, x_n|H_0)$ are the joint probability density functions of the n observations x_i under the conditions of target presence and target absence, respectively. For a linear envelope detector, the samples have a Rayleigh density under H_0 and a Ricean density under H_1, and the likelihood ratio detector reduces to

$$\prod_{i=1}^{n} I_0\left(\frac{A_i x_i}{\sigma^2}\right) \geq T \tag{7.2}$$

where I_0 is the modified Bessel function of zero order, σ^2 is the noise power, and A_i is the target amplitude of the ith pulse and is proportional to the antenna power pattern. For small signals ($A_i << \sigma$), the detector reduces to the square-law detector

$$\sum_{i=1}^{n} A_i^2 x_i^2 \geq T \tag{7.3}$$

and for large signals ($A_i >> \sigma$), it reduces to the linear detector

$$\sum_{i=1}^{n} A_i x_i > T \tag{7.4}$$

For constant signal amplitude (i.e., $A_i = A$), these detectors were first studied by Marcum[1] and were studied in succeeding years by numerous other people. The most important facts concerning these detectors are the following:

- The detection performances of the linear and square-law detectors are similar, differing only by less than 0.2 dB over wide ranges of P_D, P_{fa}, and n.
- Because the signal return of a scanning radar is modulated by the antenna pattern, to maximize the *S/N* when integrating a large number of pulses with no weighting (i.e., $A_i = 1$), only 0.84 of the pulses between the half-power points should be integrated, and the antenna beam-shape factor (ABSF) is 1.6 dB.[4] The ABSF is the number by which the midbeam *S/N* must be reduced so that the detection curves generated for equal signal amplitudes can be used for the scanning radar.
- The collapsing loss for the linear detector can be several decibels greater than the loss for a square-law detector[5] (see Figure 7.1). The collapsing loss is the additional signal required to maintain the same P_D and P_{fa} when unwanted noise samples along with the desired signal-plus-noise samples are integrated. The number of signal samples integrated is N, the number of extraneous noise samples integrated is M, and the collapsing ratio $\rho = (N + M)/N$.
- Most automatic detectors are required not only to detect targets but also to make angular estimates of the azimuth position of the target. Swerling[6] calculated the standard deviation of the optimal estimate by using the Cramer-Rao lower bound. The results

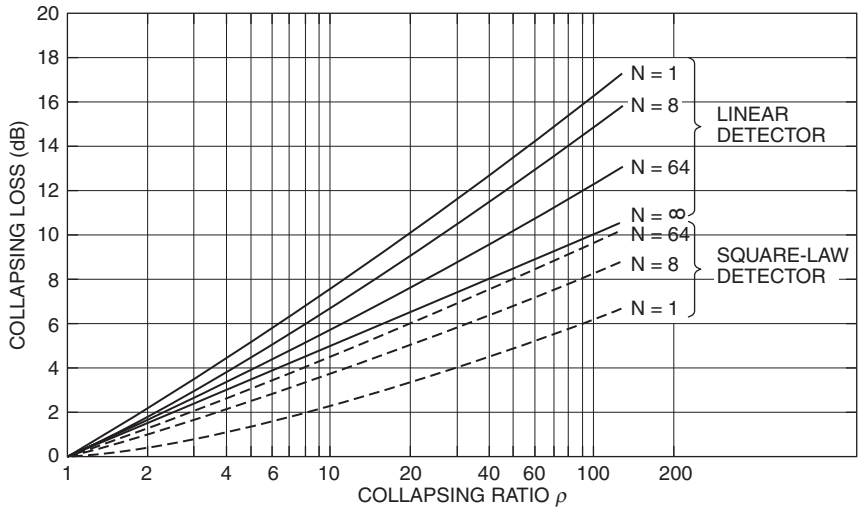


FIGURE 7.1 Collapsing loss versus collapsing ratio for a probability of false alarm of 10^{-6} and a probability of detection of 0.5 (*after G. V. Trunk[5] © IEEE 1972*)

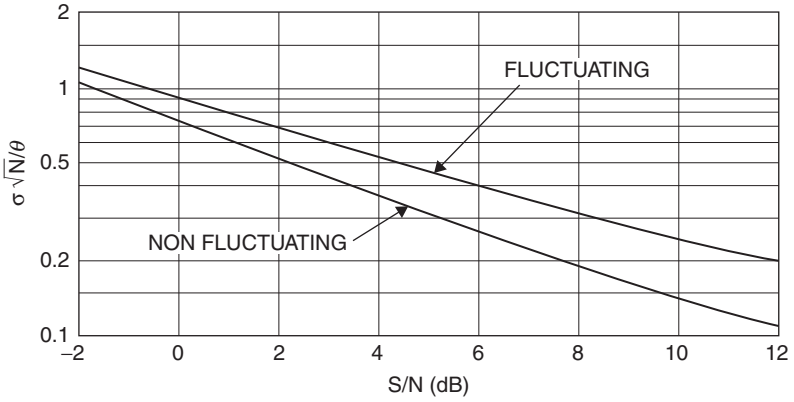


FIGURE 7.2 Cramer-Rao bound for angular estimates for fluctuating and nonfluctuating targets; σ is the standard deviation of the estimation error, and N is the number of pulses within the 3-dB beamwidth, which is θ. The *S/N* is the value at the center of the beam. (*after P. Swerling*[6] © *IEEE 1956*)

are shown in Figure 7.2, where a normalized standard deviation is plotted against the midbeam *S/N*. This result holds for a moderate or large number of pulses integrated, and the optimal estimate involves finding the location where the correlation of the returned signal and the derivative of the antenna pattern is zero. Although this estimate is rarely implemented, its performance is approached by simple estimates.

Practical Detectors. Many different detectors (often called *integrators*) are used to accumulate the radar returns as the radar sweeps by a target. A few of the most common detectors[7] are shown in Figure 7.3. The feedback integrator[8,9] and two-pole filter[10,11] are detectors that minimize the data storage requirements. While these detectors may still be found in older radars, they probably would not be implemented in new radars and will not be discussed in this edition. Though all the detectors are shown in Figure 7.3 as being constructed with shift registers, they would normally be implemented with random-access memory. The input to these detectors can be linear video, square-law video, or log video. Because linear video is probably the most commonly used, the advantages and disadvantages of the various detectors will be stated for this video.

Moving Window. The moving window in Figure 7.3*a* performs a running sum of n pulses in each range cell,

$$S_i = S_{i-1} + x_i - x_{i-n} \tag{7.5}$$

where S_i is the sum at the ith pulse of the last n pulses and x_i is the ith pulse. The performance[12] of this detector for $n \approx 10$ is only 0.5 dB worse than the optimal detector given by Eq. 7.3. The detection performance can be obtained by using an ABSF of 1.6 dB and standard detection curves for equal amplitude pulses. The angular estimate that is obtained by either taking the maximum value of the running sum or taking the midpoint between the first and last crossings of the detection threshold has a bias of $n/2$ pulses, which is easily corrected. The standard deviation of the estimation error of both these estimators is about 20 percent higher than the optimal estimate specified

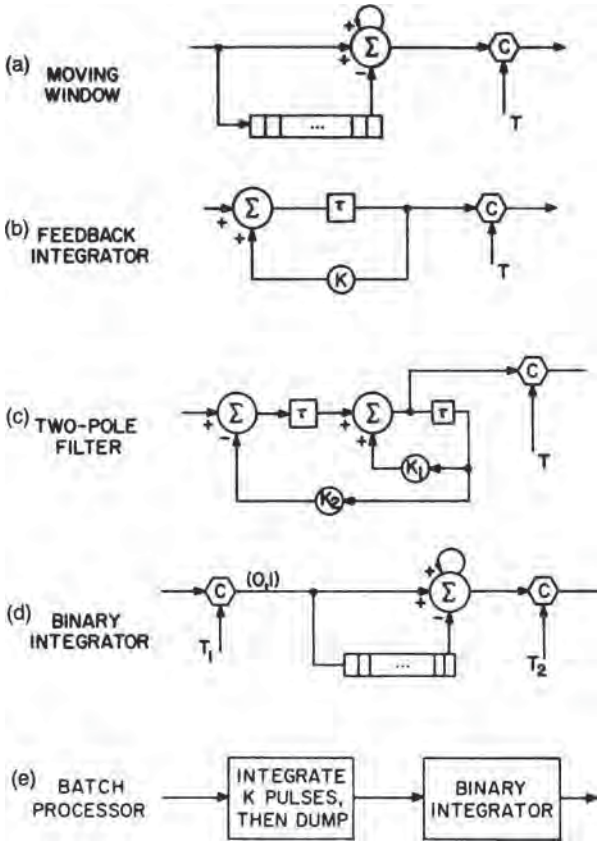


FIGURE 7.3 Block diagrams of various detectors. The letter *C* indicates a comparison, τ is a delay, and loops indicate feedback. (*from G. V. Trunk*[7])

by the Cramer-Rao bound. A disadvantage of this detector is that it is susceptible to interference; that is, one large sample from interference can cause a detection. This problem can be minimized by using soft limiting.

The detection performance discussed previously is based on the assumption that the target is centered in the moving window. In the real situation, the radar scans over the target, and decisions that are highly correlated are made at every pulse. Hansen[13] analyzed this situation for $N = 2$, 4, 8, and 16 pulses and calculated the detection thresholds shown in Figure 7.4, the detection performance shown in Figure 7.5, and the angular accuracy shown in Figure 7.6. Comparing Hansen's scanning calculation with the single-point calculation, one concludes that about 1 dB of improvement is obtained by making a decision at every pulse. The angular error of the beam-splitting procedure is about 20 percent greater than the optimal estimate. For large signal-to-noise ratios, the accuracy (rms error) of the beam-splitting and maximum-return procedures will be limited by the pulse spacing[8] and will approach

$$\sigma(\hat{\theta}) = \Delta\theta/\sqrt{12} \tag{7.6}$$

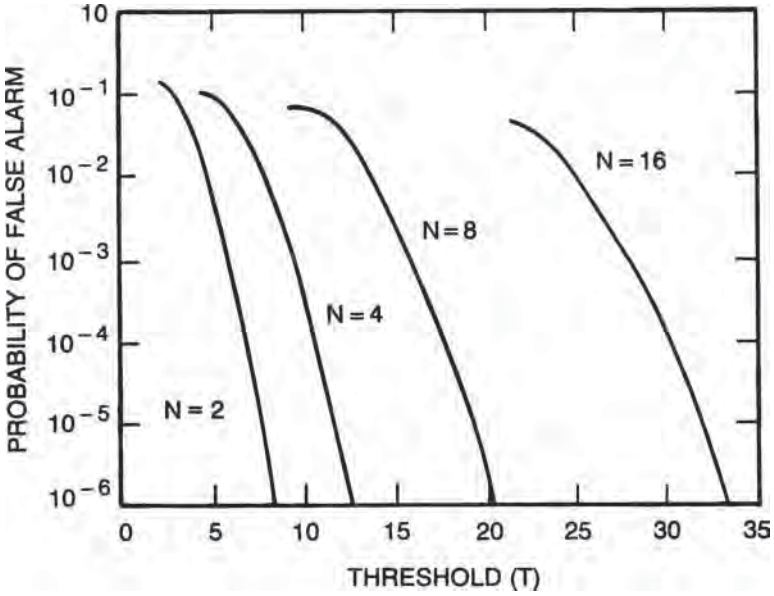


FIGURE 7.4 Single-sweep false-alarm probability P_{fa} versus threshold for moving window. The noise is Rayleigh-distributed with $\sigma = 1$. (*after V. G. Hansen*[13] *© IEEE 1970*)

where $\Delta\theta$ is the angular rotation between transmitted pulses. Consequently, if the number of pulses per beamwidth is small, the angular accuracy will be poor. For instance, if pulses are separated by 0.5 beamwidth, $\sigma(\hat{\theta})$ is bounded by 0.14 beamwidths. However, improved accuracy can be obtained by using the amplitudes of the radar returns. An accurate estimate of the target angle is given by

$$\hat{\theta} = \theta_1 + \frac{\Delta\theta}{2} + \frac{1}{2a\Delta\theta}\ln(A_2/A_1) \tag{7.7}$$

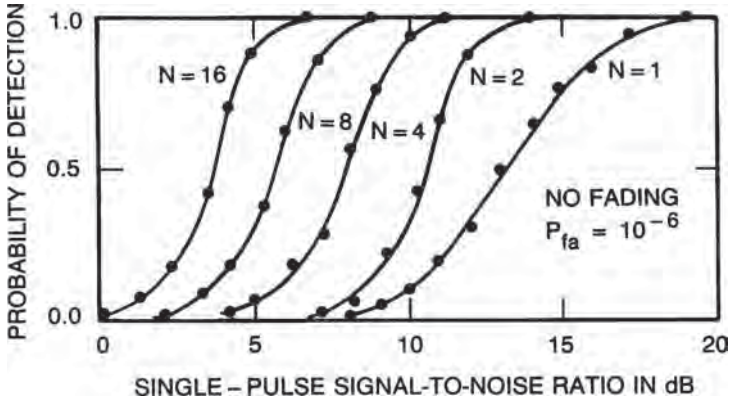


FIGURE 7.5 Detection performance of the analog moving-window detector for the no-fading case (*after V. G. Hansen*[13] *© IEEE 1970*)

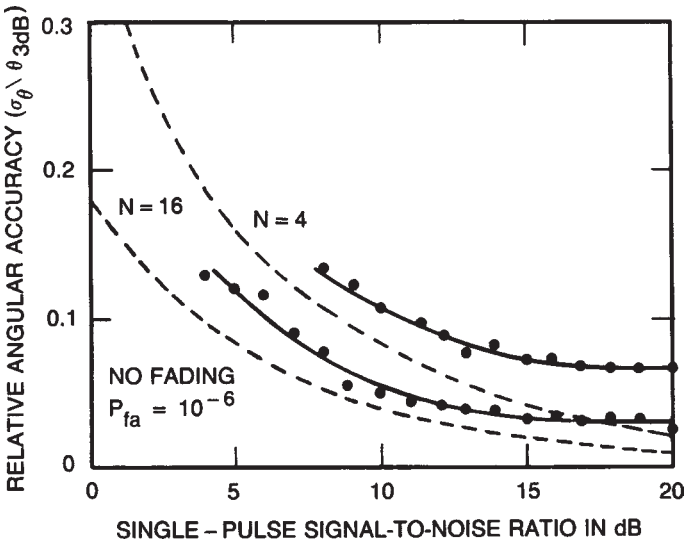


FIGURE 7.6 Angular accuracy obtained with beam-splitting estimation procedure for the no-fading case. Broken-line curves are lower bounds derived by Swerling,[6] and points shown are simulation results. (*after V. G. Hansen*[13] *© IEEE 1970*)

where

$$a = 1.386/(\text{beamwidth})^2 \tag{7.8}$$

and A_1 and A_2 are the two largest amplitudes of the returned samples and occur at angles θ_1 and $\theta_2 = \theta_1 + \Delta\theta$, respectively. Because the estimate should lie between θ_1 and θ_2 and Eq. 7.7 will not always yield such an estimate, $\hat{\theta}$ should be set equal to θ_1 if $\hat{\theta} < \theta_1$ and $\hat{\theta}$ should be equal to θ_2 if $\hat{\theta} > \theta_2$. The accuracy of this estimator is given in Figure 7.7 for the case of $n = 2$ pulses per beamwidth. This estimation procedure can also be used to estimate the elevation angle of a target in multibeam systems where θ_1 and θ_2 are the elevation-pointing angles of adjacent beams and A_1 and A_2 are the corresponding amplitudes.

Binary Integrator. The binary integrator is also known as the dual-threshold detector, M-out-of-N detector, or rank detector (see "Nonparametric Detectors," later in this section), and numerous individuals have studied it.[14–18] As shown in Figure 7.3*d*, the input samples are quantized to 0 or 1, depending on whether or not they are less than a threshold T_1. The last N zeros and ones are summed (with a moving window) and compared with a second threshold $T_2 = M$. For large N, the detection performance of this detector is approximately 2 dB less than the moving-window integrator because of the hard limiting of the data, and the angular estimation error is about 25 percent greater than the Cramer-Rao lower bound. Schwartz[16] showed that within 0.2 dB the optimal value of M for maximum P_D is given by

$$\text{M} = 1.5\sqrt{N} \tag{7.9}$$

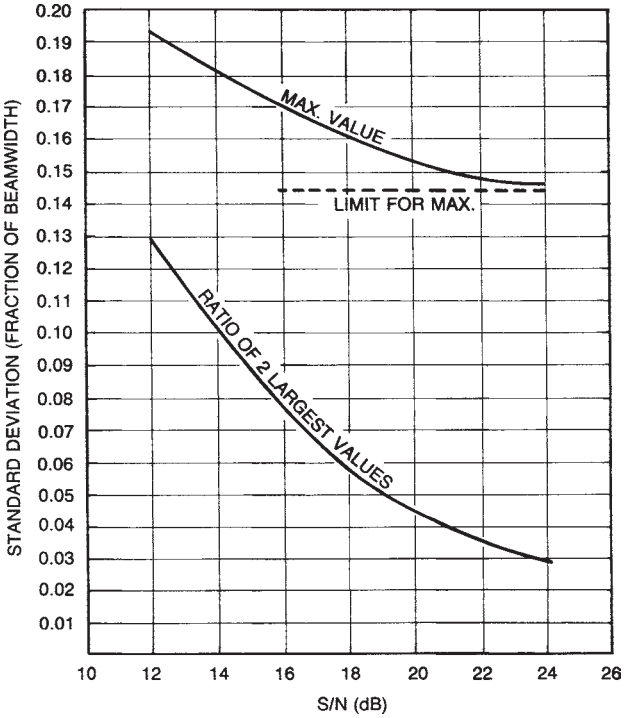


FIGURE 7.7 Angular accuracy for two pulses separated by 0.5 beamwidths

when $10^{-10} < P_{fa} < 10^{-5}$ and $0.5 < P_D < 0.9$. The optimal value of P_n, the probability of exceeding T_1 when only noise is present, was calculated by Dillard[18] and is shown in Figure 7.8. The corresponding threshold T_1 is

$$T_1 = \sigma(-2\ln P_N)^{1/2} \tag{7.10}$$

A comparison of the optimal (best value of M) binary integrator with various other procedures is given in Figures 7.9 and 7.10 for $P_D = 0.5$ and 0.9, respectively.

The binary integrator is used in many radars because (1) it is easily implemented; (2) it ignores interference spikes that cause trouble with integrators that directly use signal amplitude; and (3) it works extremely well when the noise has a non-Rayleigh density.[19] For $N = 3$, comparison of the optimal binary integrator (3 out of 3), another binary integrator (2 out of 3), and the moving-window detector in log-normal interference (an example of a non-Rayleigh density, where the log of the return has a gaussian density) is shown in Figure 7.11. The optimal binary integrator is much better than the moving-window integrator. The optimal values for log-normal interference were calculated by Schleher[19] and are $M = 3$, 8, and 25 for $N = 3$, 10, and 30, respectively.

Batch Processor. The batch processor (Figure 7.3*e*) is very useful when a large number of pulses are within the 3-dB beamwidth. If KN pulses are in the 3-dB beamwidth, K pulses are summed (batched) and either a 0 or a 1 is declared, depending on whether or not the batch is less than a threshold T_1. The last N zeros and ones are

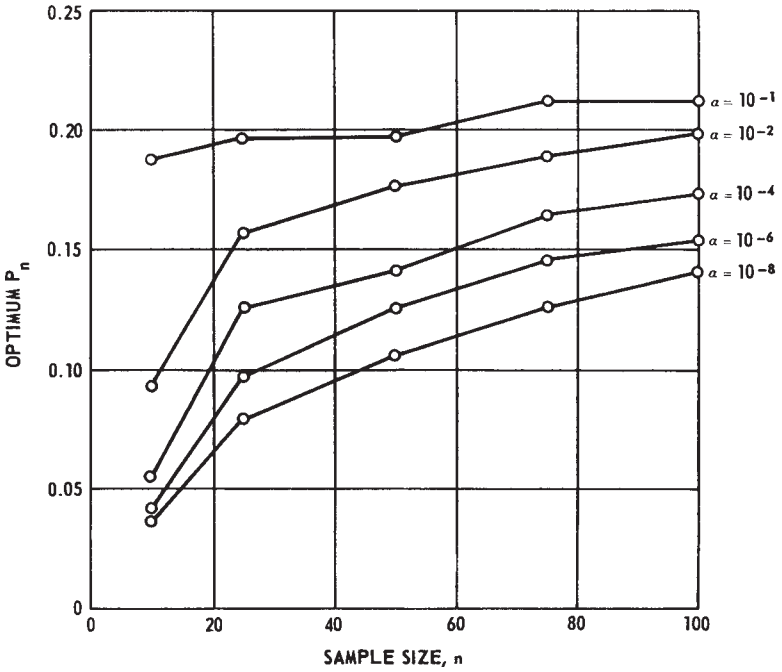


FIGURE 7.8 Optimum values of P_N as a function of the sample size n and the probability of false alarm α; Ricean distribution with $S/N = 0$ dB per pulse. (*after G. M. Dillard*[18] *© IEEE 1967*)

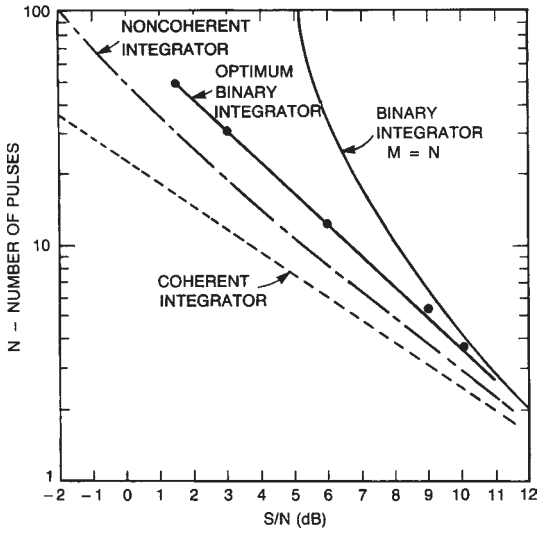


FIGURE 7.9 Comparison of binary integrator (M-out-of-N) with other integration methods ($P_{fa} = 10^{-10}$; $P_D = 0.5$) (*after M. Schwartz*[16] *© IEEE 1956*)

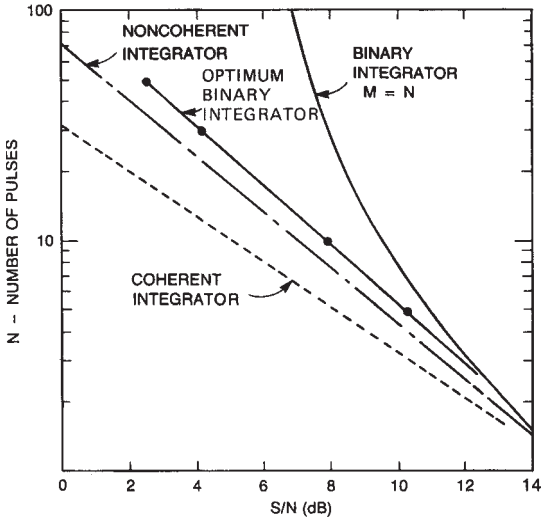


FIGURE 7.10 Comparison of binary integrator (*M*-out-of-*N*) with other integration methods ($P_{fa} = 10^{-10}$; $P_D = 0.90$) (*after M. Schwartz*[16] *© IEEE 1956*)

summed and compared with a second threshold *M*. An alternative version of this detector is to put the batch amplitudes through a moving-window detector.

The batch processor, like the binary integrator, is easily implemented, ignores interference spikes, and works extremely well when the noise has a non-Rayleigh density. Furthermore, the batch processor requires less storage, detects better, and estimates

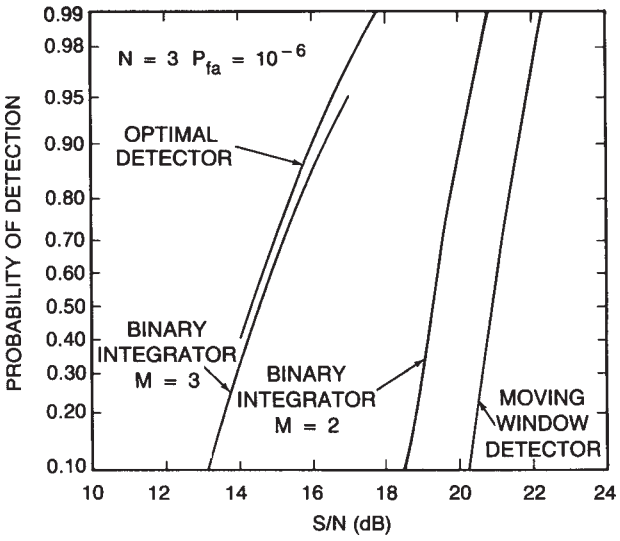


FIGURE 7.11 Comparison of various detectors in log-normal ($\sigma = 6$ dB) interference ($N = 3$; $P_{fa} = 10^{-6}$) (*after D. C. Schleher*[19] *© IEEE 1975*)

angles more accurately than the binary integrator. For instance, if there were 80 pulses on target, one could batch 16 pulses, quantize this result to a 0 or a 1, and declare a target with a 3-out-of-5 (or 2-out-of-5) binary integrator. The detection performance of the batch processor for a large number of pulses integrated is approximately 0.5 dB worse than the moving window. The batch processor has been successfully implemented by the Applied Physics Laboratory[20] of The Johns Hopkins University. To obtain an accurate azimuth estimate $\hat{\theta}$, approximately 20 percent greater than the lower bound,

$$\hat{\theta} = \frac{\sum B_i \theta_i}{\sum B_i} \tag{7.11}$$

is used, where B_i is the batch amplitude and θ_i is the azimuth angle corresponding to the center of the batch.

False-Alarm Control. In the presence of clutter, if fixed thresholds are used with the previously discussed integrators, an enormous number of detections will occur and will saturate and disrupt the tracking computer associated with the radar system. Four important facts should be noted:

- A tracking system should be associated with the automatic detection system (the only exception is when one displays multiple scans of detections).
- The P_{fa} of the detector should be matched to the tracking system to produce the overall lowest *S/N* required to form a track without initiating too many false tracks (see Figure 7.38, later in this chapter).
- Random false alarms and unwanted targets (e.g., stationary targets) are not a problem if they are removed by the tracking system.
- Scan-to-scan processing can be used to remove stationary point clutter or moving-target indication (MTI) clutter residues.

One can limit the number of false alarms with a fixed-threshold system by setting a very high threshold. Unfortunately, this would reduce target sensitivity in regions of low noise (clutter) return. Three main approaches—adaptive threshold, nonparametric detectors, and clutter maps—have been used to reduce the false-alarm problem. Adaptive thresholding and nonparametric detectors assume that the samples in the range cells surrounding the test cell (called *reference cells*) are independent and identically distributed. Furthermore, it is usually assumed that the time samples are independent. Both kinds of detectors test whether the test cell has a return sufficiently larger than the reference cells. Clutter maps allow variation in space, but the clutter must be stationary over several (typically 5 to 10) scans. Clutter maps store an average background level for each range-azimuth cell. A target is then declared in a range-azimuth cell if the new value exceeds the average background level by a specified amount.

Adaptive Thresholding. The basic assumption of the adaptive thresholding technique is that the probability density of the noise is known except for a few unknown parameters. The surrounding reference cells are then used to estimate the unknown parameters, and a threshold based on the estimated parameters is obtained. The simplest adaptive detector, shown in Figure 7.12, is the cell-average CFAR (constant false-alarm rate) investigated by Finn and Johnson.[21] If the noise has a Rayleigh density, $p(x) = x \exp(-x^2/2\sigma^2)/\sigma^2$, only the parameter σ (σ^2 is the noise power) needs to be estimated, and the threshold is of the form $T = K\Sigma x_i = Kn\sqrt{\pi/2}\hat{\sigma}$, where $\hat{\sigma}$ is the

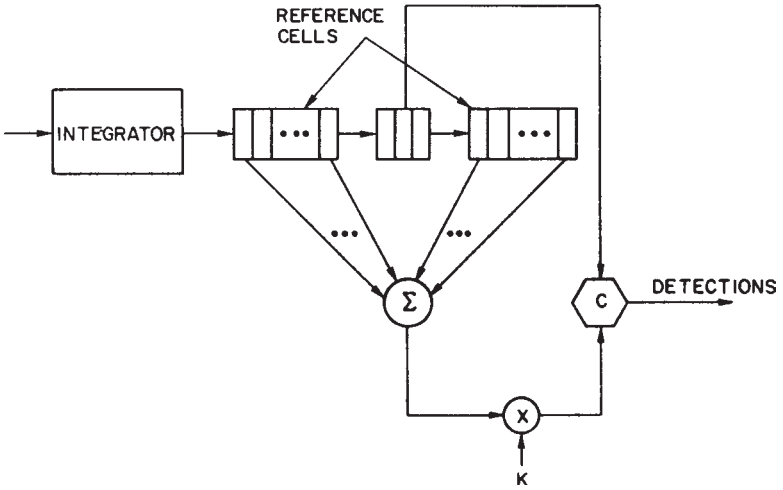


FIGURE 7.12 Cell-averaging CFAR. The letter C indicates a comparison. (*from G. V. Trunk*[7])

estimate of σ. However, since T is set by an estimate $\hat{\sigma}$, it has some error and must be slightly larger than the threshold that one would use if σ were known exactly a priori. The raised threshold causes a loss in target sensitivity and is referred to as a CFAR loss. This loss has been calculated[22] and is summarized in Table 7.1. As can be seen, for a small number of reference cells, the loss is large because of the poor estimate of σ. Consequently, one would prefer to use a large number of reference cells. However, if one does this, the homogeneity assumption (i.e., all the reference cells are statistically similar) might be violated. A good rule of thumb is to use enough reference cells so that the CFAR loss is below 1 dB and at the same time not let the reference cells extend over a range interval that violates the homogenous background assumption. Unfortunately, for a specific radar this might not be feasible.

If there is uncertainty about whether or not the noise is Rayleigh-distributed, it is better to threshold individual pulses and use a binary integrator as shown in Figure 7.13. This detector is tolerant of variations in the noise density because by setting K to yield a 1 with probability 0.1, a $P_{fa} \approx 10^{-6}$ can be obtained by using a 7-out-of-9 detector. While noise may be non-Rayleigh, it will probably be very Rayleigh-like out to the

TABLE 7.1 CFAR Loss for $P_{fa} = 10^{-6}$ and $P_D = 0.9$*

Number of Pulses Integrated	Loss for Various Numbers of Reference Cells (in dB)					
	1	2	3	5	10	∞
1	. . .	. . .	15.3	7.7	3.5	0
3	. . .	7.8	5.1	3.1	1.4	0
10	6.3	3.3	2.2	1.3	0.7	0
30	3.6	2.0	1.4	1.0	0.5	0
100	2.4	1.4	1.0	0.6	0.3	0

**(after R. L. Mitchell and J. F. Walker*[22] *© IEEE 1971)*

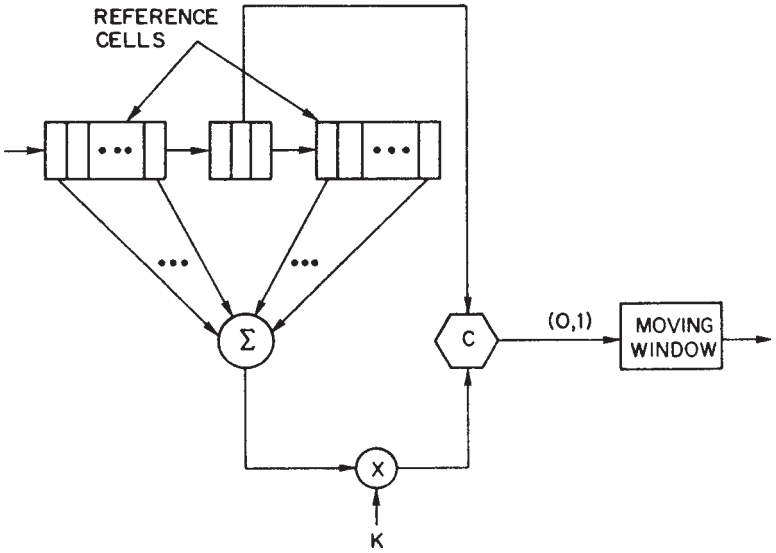


FIGURE 7.13 Implementation of a binary integrator. The letter *C* indicates a comparison. (*from G. V. Trunk*[7])

tenth percentile. Furthermore, one can use feedback based on several scans of data to control K in order to maintain a desired P_{fa} on either a scan or a sector basis. This demonstrates a general rule: to maintain a low P_{fa} in various environments, adaptive thresholding should be placed in front of the integrator.

If the noise power varies from pulse to pulse (as it would in jamming when frequency agility is employed), one must CFAR each pulse and then integrate. While the binary integrator performs this type of CFAR action, analysis[23,24] has verified that the ratio detector shown in Figure 7.14 is a better detector. The ratio detector sums signal-to-noise ratios and is specified by

$$\sum_{i=1}^{n} \frac{x_i^2(j)}{\dfrac{1}{2m}\displaystyle\sum_{k=1}^{m}\left[x_i^2(j+1+k)+x_i^2(j-1-k)\right]} \tag{7.12}$$

where $x_i(j)$ is the ith envelope-detected pulse in the jth range cell and $2m$ is the number of reference cells. The denominator is the maximum-likelihood estimate of σ_i^2, the noise power per pulse. The ratio detector will detect targets even though only a few returned pulses have a high signal-to-noise ratio. Unfortunately, this will also cause the ratio detector to declare false alarms in the presence of narrow-pulse interference. To reduce the number of false alarms when narrow-pulse interference is present, the individual power ratios can be soft-limited[24] to a small enough value so that interference will cause only a few false alarms. A comparison of the ratio detector with other commonly used detectors is shown in Figures 7.15 and 7.16 for nonfluctuating and fluctuating targets. A typical performance in sidelobe jamming when the jamming level varies by 20 dB per pulse is shown in Figure 7.17. By employing a second test to

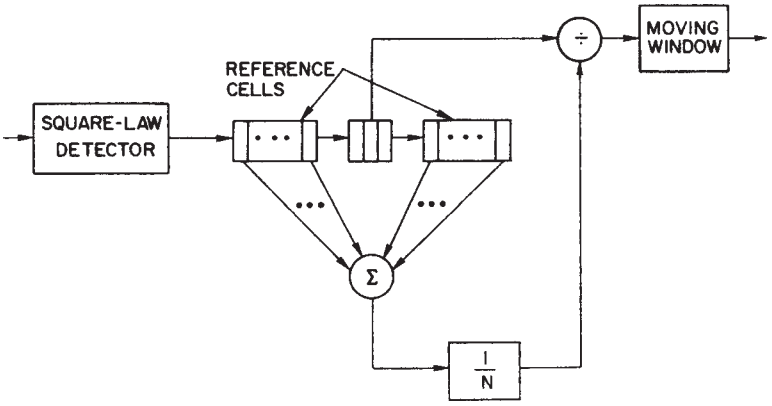


FIGURE 7.14 Ratio detector (*from G. V. Trunk*[7])

identify the presence of narrow-pulse interference, a detection performance approximately halfway between the limiting and nonlimiting ratio detectors can be obtained.

If the noise samples have a non-Rayleigh density such as the chi-square density or log-normal density, it is necessary to estimate more than one parameter and the adaptive detector is more complicated. Usually two parameters are estimated, the mean and the variance, and a threshold of the form $T = \hat{\mu} + K\hat{\sigma}$ is used. The sampled mean is easily obtained. However, the usual estimate of the standard deviation

$$\hat{\sigma} = \left[\frac{1}{N}\Sigma(x_i - \hat{\mu})^2\right]^{1/2} \tag{7.13}$$

where

$$\hat{\mu} = \frac{1}{N}\Sigma x_i \tag{7.14}$$

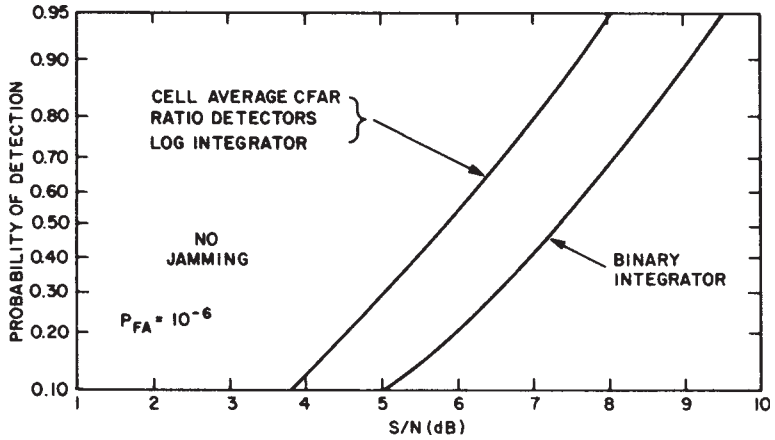


FIGURE 7.15 Curves of probability of detection versus signal-to-noise ratio per pulse for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: nonfluctuating target, $N = 6$, $2m = 16$ reference cells, and $P_{fa} = 10^{-6}$ (*from G. V. Trunk and P. K. Hughes*[24])

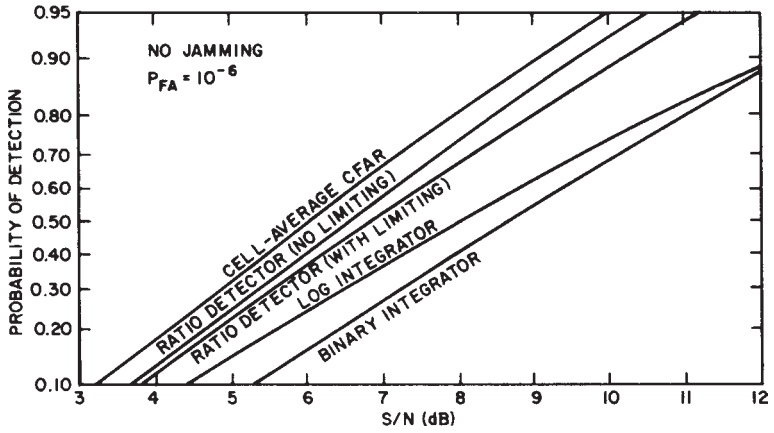


FIGURE 7.16 Curves of probability of detection versus signal-to-noise ratio for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: Rayleigh, pulse-to-pulse fluctuating target, $N = 6$, $2m = 16$ reference cells, and $P_{fa} = 10^{-6}$ (*from G. V. Trunk and P. K. Hughes*[24])

is somewhat more difficult to implement; consequently, the mean deviate defined by

$$\sigma = A \sum |x_i - \hat{\mu}| \tag{7.15}$$

is sometimes used because of its ease of implementation and because it is more robust. It should be noted that the CFAR loss associated with a two-parameter threshold is larger than those associated with a one-parameter threshold (see Table 7.1), and for that reason, a two-parameter threshold is rarely used.

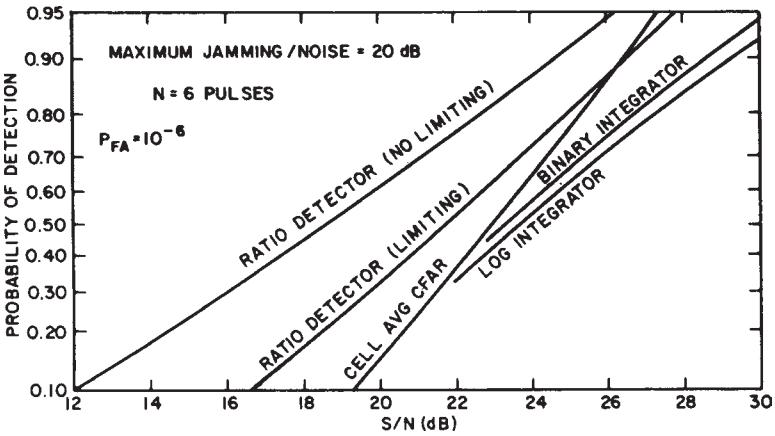


FIGURE 7.17 Curves of probability of detection versus signal-to-noise ratio for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: Rayleigh, pulse-to-pulse fluctuations, $2m = 16$ reference cells, $P_{fa} = 10^{-6}$, and maximum jamming-to-noise ratio = 20 dB (*from G. V. Trunk and P. K. Hughes*[24])

If the noise samples are correlated, nothing can be done to the binary integrator to yield a low P_{fa}. Thus, it should not be used in this situation. However, if the correlation time is less than a batching interval, the batch processor will yield a low P_{fa} without modifications.

Target Suppression. Target suppression is the loss in detectability caused by other targets or clutter residues in the reference cells. Basically, there are two approaches to solving this problem: (1) remove large return from the calculation of the threshold[25–27] or (2) diminish the effects of large returns by either limiting or using log video. The technique that should be used is a function of the particular radar system and its environment.

Rickard and Dillard[26] proposed a class of detectors D_K, where the K largest samples are censored (removed) from the reference cells. A comparison of D_0 (no censoring) with D_1 and D_2 for a Swerling 2 target and a single square-law detected pulse is shown in Figure 7.18, where N is the number of reference cells, β is the ratio of the power of the interfering target to the target in the test cell, and the bracketed pair (m, n) indicates the Swerling models of the target and the interfering target, respectively. As shown in Figure 7.18, when one has an interfering target, the P_D does not approach 1 as S/N increases. Another approach[25] that censors samples in the reference cell if they exceed a threshold is briefly discussed in the "Nonparametric Detector" subsection.

Finn[27] investigated the problem of the reference cells spanning two continuous different "noise" fields (e.g., thermal noise, sea clutter, etc.). On the basis of the samples, he estimated the statistical parameters of the two noise fields and the separation point between them. Then, only those reference cells that are in the noise field containing the test cell are used to calculate the adaptive threshold.

An alternative approach for interfering targets is to use log video. By taking the log, large samples in the reference cells will have less effect than linear video on the threshold. The loss associated with using log video, rather than linear video, is 0.5 dB

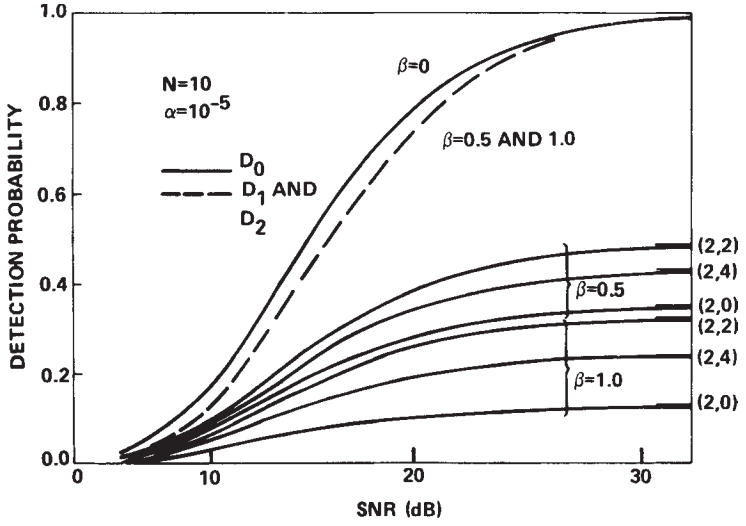


FIGURE 7.18 Detection probability versus SNR for a Swerling Case 2 primary target (*after J. T. Rickard and G. M. Dillard*[26] *© IEEE 1977*)

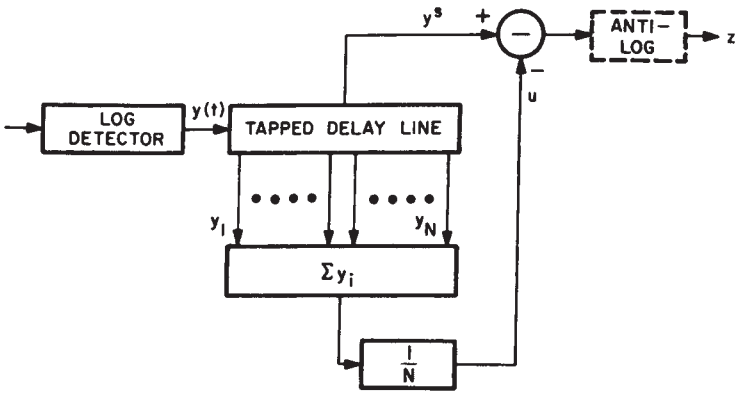


FIGURE 7.19 Block diagram of cell-averaging log-CFAR receiver (*after V. G. Hansen and J. R. Ward*[29] *© IEEE 1972*)

for 10 pulses integrated and 1.0 dB for 100 pulses integrated.[28] An implementation of the log CFAR[29] is shown in Figure 7.19. In many systems, the antilog shown in Figure 7.19 is not taken. To maintain the same CFAR loss as for linear video, the number of reference cells $M_{\log}$ for the log CFAR should equal

$$M_{\log} = 1.65M_{\text{lin}} - 0.65 \tag{7.16}$$

where M_{lin} is the number of reference cells for linear video. The effect of target suppression with log video is discussed later in this section (see Table 7.2, later in the chapter).

Nonparametric Detectors. Usually nonparametric detectors obtain CFAR by ranking the test sample with the reference cells.[30,31] Ranking means that one orders the samples from the smallest to the largest and replaces the smallest with rank 0, the next smallest with rank 1, . . . , and the largest with rank $n - 1$. Under the hypothesis that all the samples are independent samples from an unknown density function, the test sample has equal probability of taking on any of the n values. For instance, referring to the ranker in Figure 7.20, the test cell is compared with 15 of its neighbors. Since, in the set of 16 samples, the test sample has equal probability of being the smallest sample (or equivalently any other rank), the probability that the test sample takes on values 0, 1, . . . , 15 is 1/16. A simple rank detector is constructed by comparing the rank with a threshold K and generating a 1 if the rank is larger, a 0 otherwise. The 0s and 1s are summed in a moving window. This detector incurs a CFAR loss of about 2 dB but achieves a fixed P_{fa} for any unknown noise density as long as the time samples are independent. This detector was incorporated into the ARTS-3A postprocessor used in conjunction with the Federal Aviation Administration airport surveillance radar (ASR). The major shortcoming of this detector is that it is fairly susceptible to target suppression (e.g., if a large target is in the reference cells, the test cell cannot receive the highest ranks).

If the time samples are correlated, the rank detector will not yield CFAR. A modified rank detector, called the modified generalized sign test (MGST),[25] maintains a low P_{fa} and is shown in Figure 7.21. This detector can be divided into three parts: a ranker, an integrator (in this case, a two-pole filter), and a threshold (decision process). A target is declared when the integrated output exceeds two thresholds.

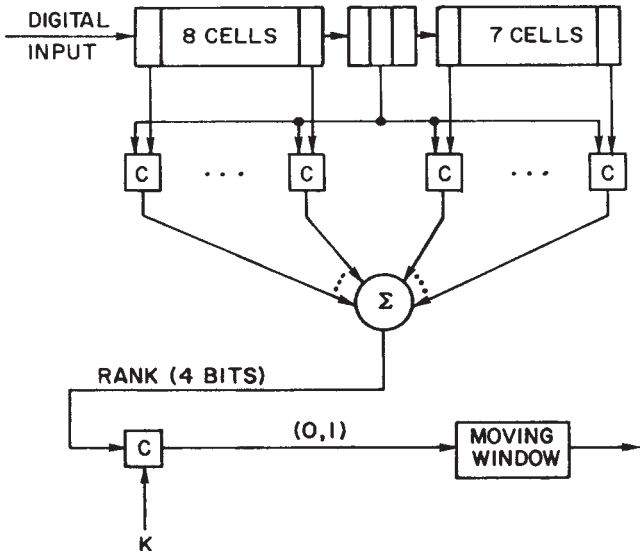


FIGURE 7.20 Rank detector: output of a comparator C is either a zero or a one (*from G. V. Trunk*[7])

The first threshold is fixed (equals $\mu + T_1/K$ in Figure 7.21) and yields $P_{fa} = 10^{-6}$ when the reference cells are independent and identically distributed. The second threshold is adaptive and maintains a low P_{fa} when the reference samples are correlated. The device estimates the standard deviation of the correlated samples with the mean deviate estimator, where extraneous targets in the reference cells have been excluded from the estimate by use of a preliminary threshold T_2.

The basic disadvantages of all nonparametric detectors are that (1) they have relatively large CFAR losses; (2) they have problems with correlated samples; and (3) one loses amplitude information, which can be a very important discriminant between target and clutter.[32] For example, a large return (*cross section* ≥ 1000 m^2) in a clutter area is probably just clutter breakthrough. See "Radar Detection Acceptance" in Section 7.3.

Clutter Mapping. A clutter map uses adaptive thresholding where the threshold is calculated from the return in the test cell on previous scans rather than from the surrounding reference cells on the same scan. This technique has the advantage in that for essentially stationary environments (e.g., land-based radar against ground clutter), the radar has interclutter visibility—it can see between large clutter returns. Lincoln Laboratory[33] in its moving-target detector (MTD) used a clutter map for the zero-doppler filter very effectively. The decision threshold T for the ith cell is

$$T = A\ S_{i-1} \tag{7.17}$$

where the clutter is estimated using a simple feedback integrator

$$S_i = K\ S_{i-1} + X_i \tag{7.18}$$

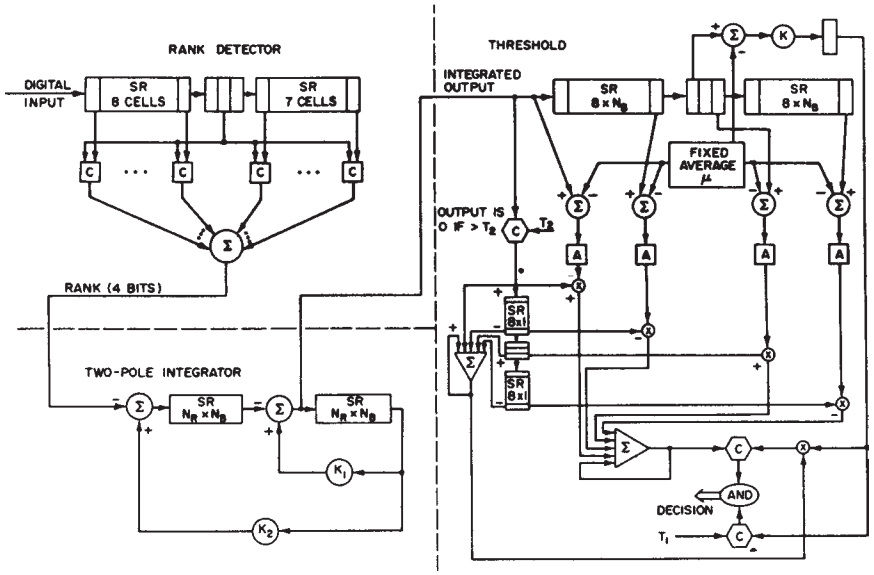


FIGURE 7.21 Modified generalized sign test processor (*after G. V. Trunk et al.*[25] *1974*)

where S_i is the average background level, X_i is the return in the ith cell, K is the feedback value that determines the map time constant, and A is the constant that determines the P_{fa}. In the MTD used for the ASR application, K is 7/8, which effectively averages the last eight scans. The purpose of the clutter map is to detect, in clutter free areas, crossing targets that would have been removed by the doppler processing. The main utility of clutter maps is with fixed-frequency, land-based radars. While clutter maps can be used with frequency-agile radars and on moving platforms (e.g., radars on ships), they are not nearly as effective in these environments.

Target Resolution. In automatic detection systems, a single large target will probably be detected (i.e., cross a detection threshold) many times, e.g., in adjacent range cells, azimuth beams, and elevation beams. Therefore, automatic detection systems have algorithms for merging the individual detections into a single centroided detection. Most algorithms have been designed so that they will rarely split a single target into two targets. This procedure results in poor range resolution capability. A merging algorithm[34] often used is the adjacent-detection merging algorithm, which decides whether a new detection is adjacent to any of the previously determined sets of adjacent detections. If the new detection is adjacent to any detection in the set of adjacent detections, it is added to the set. Two detections are adjacent if two of their three parameters (range, azimuth, and elevation) are the same and the other parameter differs by the resolution element: range cell ΔR, azimuth beamwidth θ, or elevation beamwidth γ.

A study[34] compared the resolving capability of three common detection procedures: linear detector with $T = \hat{\mu} + A\hat{\sigma}$, linear detector with $T = B\hat{\mu}$, and log detector with $T = C + \hat{\mu}$, where the constants A, B, and C are used to obtain the same P_{fa} for all detectors. The estimates $\hat{\mu}$ and $\hat{\sigma}$ of μ and σ were obtained from either (1) all the reference cells or (2) the leading or lagging half of the reference cells, choosing the

TABLE 7.2 Probability of Detecting Both Targets with Log Video When the Two Targets Are Separated by 1.5, 2.0, 2.5, or 3.0 Range Cells (*S/N* of target 1 is 20 dB and *S/N* of target 2 is 10, 13, 20, 30, or 40 dB)*

Thresholding Technique	Target Separation	S/N of Target no. 2				
		10	13	20	30	40
All reference cells	1.5	0.00	0.04	0.00	0.00	0.00
	2.0	0.00	0.22	0.54	0.14	0.10
	2.5	0.04	0.24	0.94	0.62	0.32
	3.0	0.00	0.24	0.88	0.92	0.78
Reference cells with minimum mean value	1.5	0.00	0.00	0.00	0.00	0.02
	2.0	0.10	0.32	0.44	0.12	0.04
	2.5	0.18	0.58	0.98	0.46	0.28
	3.0	0.22	0.66	0.98	0.82	0.74

*(*after G. V. Trunk*[34] *© IEEE 1978*)

half with the lower mean value. The first simulation involved two targets separated by 1.5, 2.0, 2.5, or 3.0 range cells and a third target 7.0 range cells from the first target. When the two closely spaced targets were well separated, either 2.5 or 3.0 range cells apart, the probability of detecting both targets (P_{D2}) was < 0.05 for the linear detector with $T = \hat{\mu} + A\hat{\sigma}$; $0.15 < P_{D2} < 0.75$ for the linear detector with $T = B\hat{\mu}$; and $P_{D2} > 0.9$ for the log detector. A second simulation, involving only two targets, investigated the effect of target suppression on log video, and the results are summarized in Table 7.2. The maximum value of P_{D2} is obtained when both targets have an *S/N* of 20 dB. If one of the targets has a larger *S/N* than the other target, suppression occurs—either target 1 suppresses target 2 or vice versa. Also, one notes an improved performance for a small *S/N* (10 to 13 dB) when calculating the threshold using only the half of the reference cells with the lower mean value. The resolution capability of the log detector that uses only the half of the reference cells with the lower mean is shown in Figure 7.22. The probability of resolving two, equal-amplitude targets does not rise above 0.9 until they are separated in range by 2.5 pulse widths.

By assuming that the target is small with respect to the pulse width and that the pulse shape is known, the resolution capability can be improved by fitting the known pulse shape to the received data and comparing the residue square error with a threshold.[35] If only one target is present, the residue should be only noise and hence should be small. If two or more targets are present, the residue will contain signal from the remaining targets and should be large. The results of resolving two targets with *S/N* = 20 dB are shown in Figure 7.23. These targets can be resolved at a resolution probability of 0.9 with a false alarm probability of 0.01 at separations varying between one-fourth and three-fourths of a pulse width, depending on the relative phase difference between the two targets. Moreover, this result can be improved further by processing multiple pulses.

Automatic Detection Summary. When only 2 to 4 samples (pulses) are available, a binary integrator should be used to avoid false alarms due to interference. When a moderate number of pulses (5 to 16) are available, a binary integrator or a moving-window integrator should be used. If the number of pulses is large (greater than 20), a batch processor should be used. If the samples are independent, a one-parameter (mean) threshold can be used. If the samples are dependent, one can either use

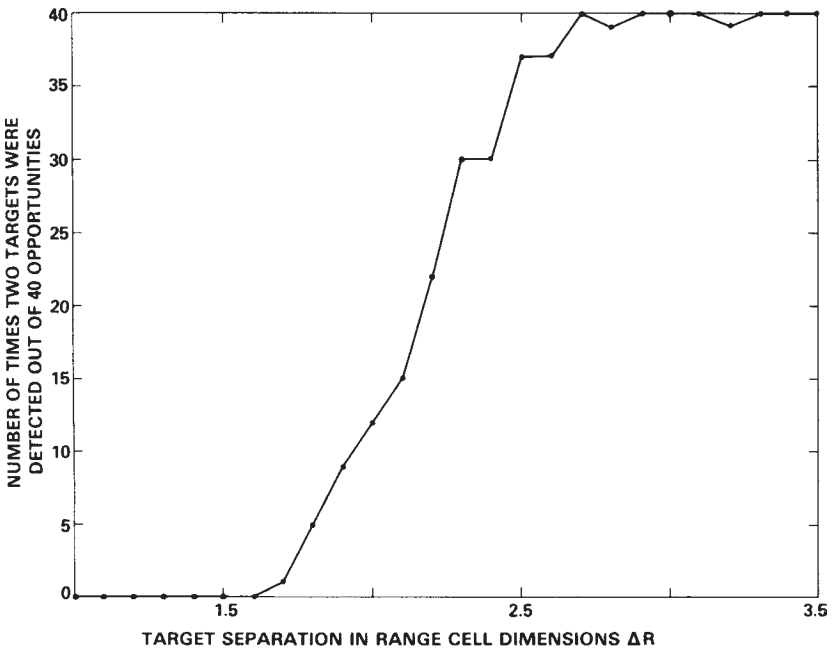


FIGURE 7.22 Resolution capability of a log detector that used half of the references cells with lower mean (*after G. V. Trunk*[34] *© IEEE 1978*)

a two-parameter (mean and variance) threshold or adapt a one-parameter threshold on a sector basis. However, these rules should serve only as a general guideline. It is *highly recommended* that before a detector is chosen the radar video from the environment of interest be collected and analyzed and that various detection processes be simulated on a computer and tested against the recorded data.

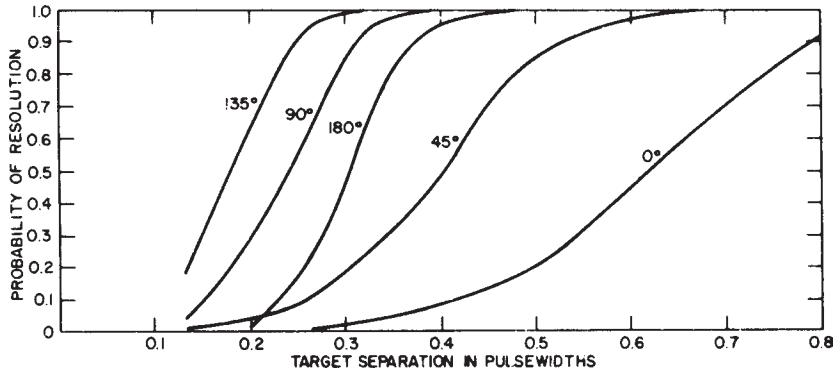


FIGURE 7.23 Probability of resolution as a function of range separation: probability of false alarm is 0.01; sampling rate $\Delta R = 1.5$ samples per pulse width; target strengths, nonfluctuating, A1 = A2 = 20 dB; phase differences = 0°, 45°, 90°, 135°, and 180°. (*after G. V. Trunk*[35] *© IEEE 1984*)

Many modern radars use coherent processing to remove clutter. For the purpose of applying the previous discussions on noncoherent processing to coherent processing, the integrated output in a range-doppler cell of the doppler processor for a single coherent processing interval (CPI) can be treated as a single noncoherent pulse. Because three ambiguous measurements (i.e., detections) are usually required to remove the range and doppler ambiguities,[36,37] 4 to 8 CPIs may be transmitted, and hence, there are usually 4 to 8 noncoherent pulses available for processing.

7.3 AUTOMATIC TRACKING

A track represents the belief that a physical object or "target" is present and has actually been detected by the radar. An automatic radar tracking system forms a track when enough radar detections are made in a believable enough pattern to indicate a target is actually present (as opposed to a succession of false alarms) and when enough time has passed to allow accurate calculation of the target's kinematic state—usually position and velocity. Thus, the goal of tracking is to transform a (time-lapse) detection picture (shown in Figure 7.24*a*), consisting of target detections, false alarms, and clutter, into a track picture (shown in Figure 7.24*b*), consisting of tracks on real targets, occasional false tracks, and occasional deviations of track position from true target positions.

Figures 7.24*a* and 7.24*b* also illustrate some of the challenges of automatic tracking. Detections are made on targets, but some detections are missing because of target fades or multiple targets in the same resolution cell, whereas additional detections are present due to clutter or noise.

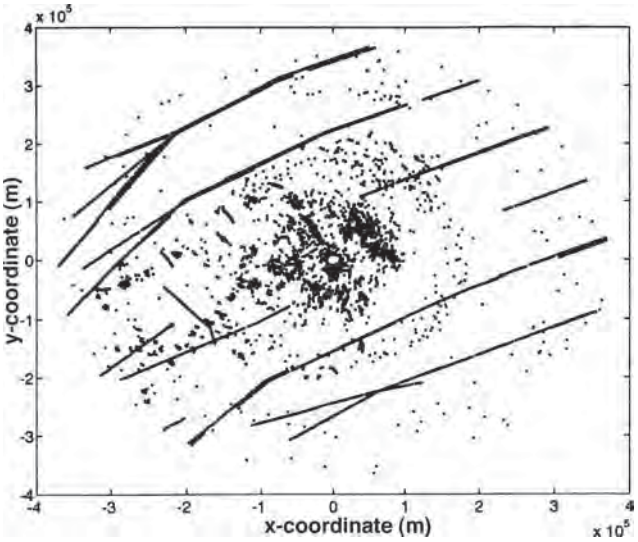


FIGURE 7.24*a* Thirty-minute time lapse of AN/FPN-504 (L band) air traffic control radar detections over a ±400-km square area (*after H. Leung et al.*[38] *© IEEE 1999*)

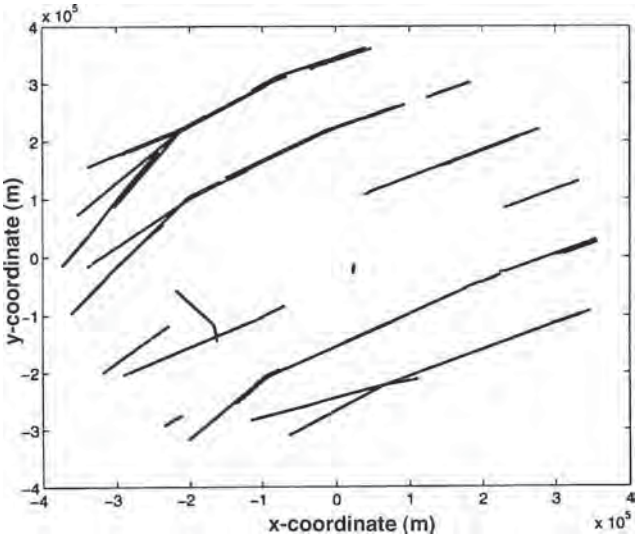


FIGURE 7.24*b* Thirty-minute time lapse of tracks formed from data in Figure 7.24*a,* using Global Nearest-Neighbor (GNN) Technique (*after H. Leung et al.*[38] *© IEEE 1999*)

Automatic tracking can generally be divided into the five steps shown in Figure 7.25 and detailed here:

1. Radar detection acceptance: accepting or rejecting detections for insertion into the tracking process. The purpose of this step is to control false track rates.
2. Association of accepted detections with existing tracks.
3. Updating existing tracks with associated detections.
4. New track formation using unassociated detections.
5. Radar scheduling and control.

The result of the automatic tracking process is a track file that contains a track state for each target detected by the radar.

As shown in Figure 7.25, there is a feedback loop between all these functions so the ability to update existing tracks accurately naturally affects the ability to associate detections with existing tracks. Also, the ability to correctly associate detections with existing tracks affects the track's accuracy and the ability to correctly distinguish between an existing track and a new one. The detection accept/reject step makes use of feedback from the association function that measures the detection activity in different regions of the radar coverage. More stringent acceptance criteria are applied in more active regions.

Track File. When a track is established in the computer, it is assigned a track number. All parameters associated with a given track are referred to by this track number. Typical track parameters are the filtered and predicted position; velocity; acceleration (when applicable); time of last update; track quality; signal-to-noise ratio; covariance matrices (the covariance contains the accuracy of all the track coordinates and all the statistical cross-correlations between them), if a Kalman-type filter is being used; and

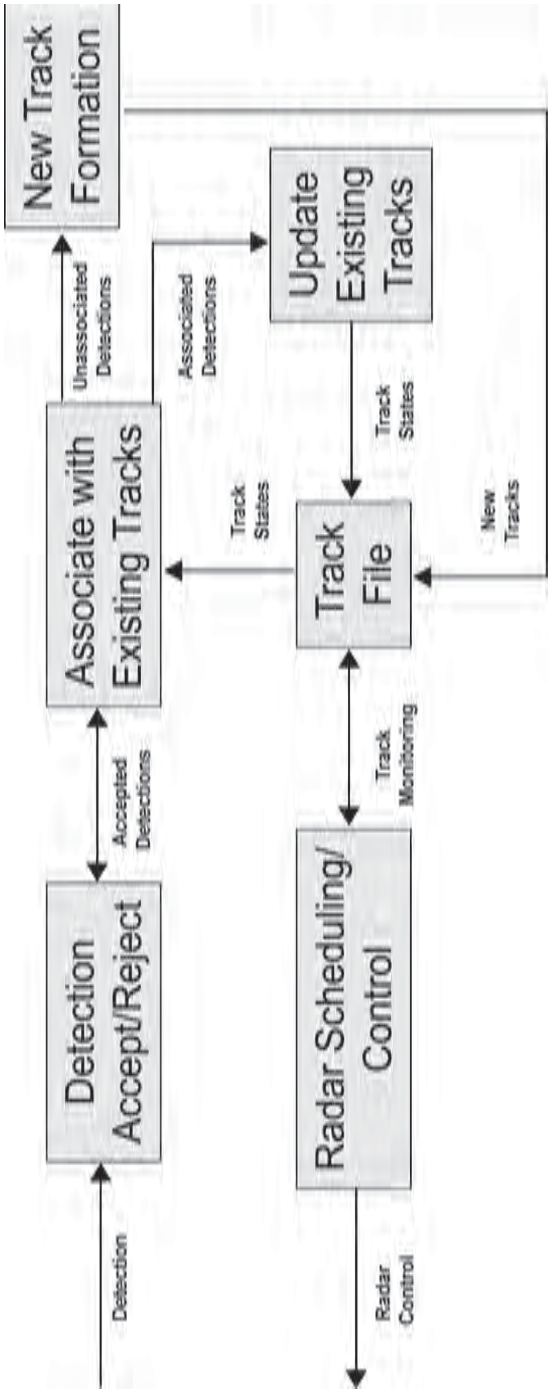


FIGURE 7.25 Structure of automatic tracking process

track history (i.e., the last *n* detections). Tracks and detections can be accessed in various sectored, linked-list, and other data structures so that the association process can be performed efficiently.[39] In addition to the track file, a clutter file is maintained. A clutter number is assigned to each stationary or very slowly moving echo. All parameters associated with a clutter point are referred to by this clutter number. Again, each clutter number is assigned to a sector in azimuth for efficient association.

Radar Detection Acceptance. When the radar system has either no or limited coherent processing, not all the detections declared by the automatic detector are used in the tracking process. Rather, many of the detections (contacts) are filtered out in software using a process called *activity control.*[32,40] The basic idea is to use detection signal characteristics in connection with a map of the detection activity to reduce the rate of detections to one that is acceptable for forming tracks. The map is constructed by counting the unassociated detections (those that do not associate with existing tracks) at the point in the track processing shown in Figure 7.25.

Counts are averaged over many revisits of the radar to achieve statistical significance. The detection signal characteristics (such as amplitude or signal-to-noise) are then re-thresholded to reduce sensitivity in regions of unacceptably high activity. In no circumstances are detections eliminated if they fall within a track gate (i.e., a gate centered on the predicted position of a firm track). Figure 7.26 illustrates an example

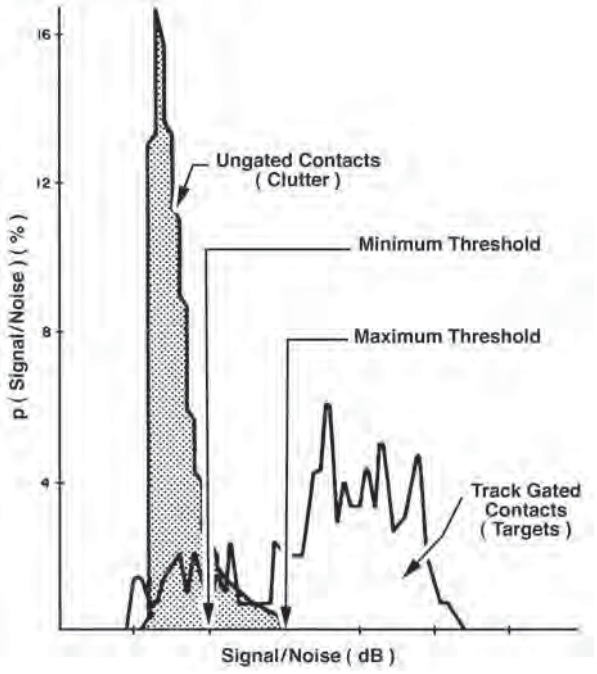


FIGURE 7.26 Histogram of detection signal-to noise ratio detection illustrating the effectiveness of the activity control using the signal-to-noise test in rain clutter. Ungated contacts generally represent clutter. Gated contacts generally represent targets. Re-thresholding, in this case, successfully eliminates large numbers of clutter detections while preserving most target detections. (*after W. G. Bath et al.*[32])

of this process when large numbers of rain clutter detections are potentially overloading the tracking process. In this case, activity control effectively eliminates most of the clutter detections without eliminating many of the actual target detections. However, because this process essentially constitutes controlled desensitization of the radar, it must be used with care. The mapping of the detection activity must be precise so that desensitization occurs only in those regions requiring it.

Updating Existing Tracks with Associated Detections. The simplest method of updating a track state is the α-β filter[41] described by

$$x_s(k) = x_p(k) + \alpha[x_m(k) - x_p(k)] \tag{7.19}$$

$$v_s(k) = v_s(k-1) + \beta[x_m(k) - x_p(k)]/T \tag{7.20}$$

$$x_p(k+1) = x_s(k) + v_s(k)T \tag{7.21}$$

where $x_s(k)$ is the filtered position, $v_s(k)$ is the filtered velocity, $x_p(k)$ is the predicted position, $x_m(k)$ is the measured position, T is the time between detections, and (α, β) are the position and velocity gains, respectively. The selection of (α, β) is a design tradeoff. Small gains make a small correction in the direction of each detection. As a result, the tracking filter is less sensitive to noise but is more sluggish to respond to maneuvers—deviation from the assumed target model. Conversely, large gains produce more tracking noise but quicker response to maneuvers. These errors are readily calculated as a function of α and β using the formulas shown in Table 7.3.

To tune the α-β filter for radar tracking, one uses the radar parameters to calculate the tracking errors listed in Table 7.3 as a function of the tracking gains α and β. Then one selects the gains that best meet the needs of the application. For example, consider a radar that has 50-meter range measurement accuracy and a two-second constant update interval. The application of this radar system is to track a target that moves linearly but with occasional unpredictable maneuvers of up to 1 g (9.8 m/s^2).

TABLE 7.3 Characterization of Tracking Errors as a Function of Tracking Gains α and β

Error Source	Steady-state Track Error	In Position	In Velocity
Radar detection noise (standard deviation- σ)	Standard deviation of filtered tracking state	$\sigma\left[\frac{2\alpha^2+\beta(2-3\alpha)}{\alpha([4-2\alpha-\beta])}\right]^{1/2}$	$\frac{\sigma}{T}\times\left[\frac{2\beta^2}{\alpha[4-2\alpha-\beta]}\right]^{1/2}$
Radar detection noise (standard deviation- σ)	Standard deviation of predicted tracking state	$\sigma\left[\frac{2\alpha^2+\alpha\beta+2\beta}{\alpha([4-2\alpha-\beta])}\right]^{1/2}$	$\frac{\sigma}{T}\times\left[\frac{2\beta^2}{\alpha[4-2\alpha-\beta]}\right]^{1/2}$
Constant maneuver—a (units of g's)	Lag (bias) in filtered track state	$aT^2\frac{(1-\alpha)}{\beta}$	$aT\left(\frac{\alpha}{\beta}-\frac{1}{2}\right)$
Constant maneuver—a (units of g's)	Lag (bias) in predicted track state	$\frac{aT^2}{\beta}$	$aT\left(\frac{\alpha}{\beta}+\frac{1}{2}\right)$

For simplicity, assume the Benedict-Bordner constant relationship[41] [$\beta = \alpha^2/(2 - \alpha)$] between α and β.

The position accuracy of the filter can then be calculated using the formulas in Table 7.3 and is shown in Figure 7.27. When the target is nonmaneuvering, accuracy, as measured by the standard deviation of the predicted tracking state, improves monotonically as the tracking gain α decreases to 0. Conversely, when the target is performing the 1-g maneuver, accuracy, as measured by the lag (or bias) in the predicted tracking state, improves monotonically as the tracking gain increases to 1. The total tracking error can be defined as the error that is exceeded only 1% of the time due to the sum of random errors and bias. The total range-tracking error is best in the region $0.6 < \alpha < 0.9$ with a minimum around 0.75. If accuracy for maneuvers is the dominant concern, then one would probably tune this filter to 0.75 to achieve the lowest total error for a 1-g acceleration. This same technique can be applied to many different radar-tracking problems using the equations in Table 7.3 to calculate a graph such as the one shown in Figure 7.27.

For simple tracking problems, the α-β filter with constant gains selected for the application will often be adequate. However, more complex tracking problems require variable tracking gains (e.g., larger gains at the beginning of the track and larger gains after missed detections or when the range to the track decreases, making angle noise less of an issue). A systematic method for calculating the gains depending on the situation is the

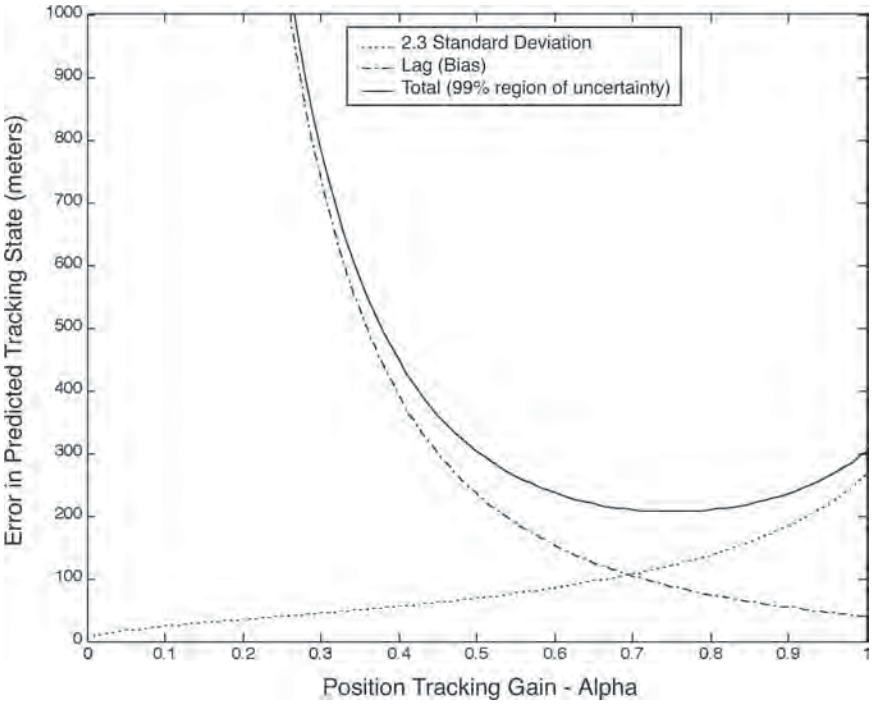


FIGURE 7.27 Example of the tuning of an α-β radar range-tracking filter by selecting the gain that minimizes total error (radar parameters: range accuracy, 50 meters; update interval, 2 seconds; target parameter, 1 g unknown acceleration; gain relation, [$\beta = \alpha^2/(2 - \alpha)$])

Kalman filter.[42,43] The Kalman filter minimizes the mean-square prediction error when the random processes are gaussian. The Kalman filter can be formulated for target motion in one, two, or three dimensions in polar, Cartesian, or Earth-centered coordinates and for three-dimensional, two-dimensional, or one-dimensional radar measurements. For simplicity, a three-dimensional tracking problem in Cartesian space with three measured radar dimensions is considered here. Target motion is described by

$$X(t_{k+1}) = \phi(t_k)\, X(t_k) + A(t_k) + A_p(t_k) \tag{7.22}$$

where $X(t_k)$ is the target state at time t_k, consisting of position and velocity components; $\phi(t_k)$ is a transition matrix that moves the target linearly over an elapsed time, $T_k = t_{k+1} - t_k$, from time t_k to time t_{k+1}; $A(t_k)$ is the target state change due to an unknown acceleration caused by a maneuver or atmospheric drag; and $A_p(t_k)$ is target state change due to a known acceleration that can be corrected, such as gravity for a falling object or Coriollis acceleration. The components of the state vector and transition matrix for this problem are[44]

$$X(t_k) = \begin{vmatrix} x(t_k) \\ \dot{x}(t_k) \\ y(t_k) \\ \dot{y}(t_k) \\ z(t_k) \\ \dot{z}(t_k) \end{vmatrix} \quad \phi(t_k) = \begin{vmatrix} 1 & T_k & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & T_k & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & T_k \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} \tag{7.23}$$

The unknown acceleration $A(t_k)$ is zero-mean and is characterized by its covariance matrix $Q(t_k)$. If one views the unknown maneuver as a white-noise process with spectral density q g/Hz, then the acceleration is sampled by each radar detection producing a discrete covariance matrix:

$$Q(t_k) = q \begin{vmatrix} T_k^3/3 & T_k^2/2 & 0 & 0 & 0 & 0 \\ T_k^2/2 & T_k & 0 & 0 & 0 & 0 \\ 0 & 0 & T_k^3/3 & T_k^2/2 & 0 & 0 \\ 0 & 0 & T_k^2/2 & T_k & 0 & 0 \\ 0 & 0 & 0 & 0 & T_k^3/3 & T_k^2/2 \\ 0 & 0 & 0 & 0 & T_k^2/2 & T_k \end{vmatrix} \tag{7.24}$$

The observation equation relates the actual radar measurements Y_k at time t_k to the target state

$$Y_k = h(X(t_k)) + n_k \tag{7.25}$$

where n_k is the radar measurement noise having a covariance matrix

$$\Re_k = \begin{vmatrix} \sigma_r^2 & 0 & 0 & 0 \\ 0 & \sigma_\theta^2 & 0 & 0 \\ 0 & 0 & \sigma_\varphi^2 & 0 \\ 0 & 0 & 0 & \sigma_D^2 \end{vmatrix} \tag{7.26}$$

composed of the radar measurement accuracies in range, azimuth, elevation, and doppler. The function h is the coordinate transform that relates the measurements to the state at time t_k according to the coordinate frame design choices (see Table 7.5, later in the chapter). In order to use the Kalman filter, h is approximated as a linear function in the vicinity of the predicted track state

$$h(X) = h(\hat{X}(t_{k+1} \mid t_k) + H[X - \hat{X}(t_{k+1} \mid t_k)] \times (X - \hat{X}(t_{k+1} \mid t_k) \tag{7.27}$$

where H is the gradient of h. Each coordinate frame has its own approximation for H. For example, if the state coordinate system is composed of three-dimensional Cartesian coordinates centered at the radar, then multiplication by H transforms Cartesian coordinates (x, y, z) into polar measurement coordinates (range, azimuth, elevation, doppler) and

$$H = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} & \frac{z}{r} & 0 & 0 & 0 \\ \frac{y}{x^2+y^2} & \frac{-x}{x^2+y^2} & 0 & 0 & 0 & 0 \\ \frac{-xz}{r^2\sqrt{x^2+y^2}} & \frac{-yz}{r^2\sqrt{x^2+y^2}} & \frac{\sqrt{x^2+y^2}}{r^2} & 0 & 0 & 0 \\ \frac{\dot{x}r - x\dot{r}}{r^2} & \frac{\dot{y}r - y\dot{r}}{r^2} & \frac{\dot{z}r - z\dot{r}}{r^2} & \frac{x}{r} & \frac{y}{r} & \frac{z}{r} \end{bmatrix} \tag{7.28}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is range.

The Kalman filter equations for radar tracking are then, simply, generalizations of the $\alpha - \beta$ filter equations where α and β vary with time. The Kalman filter update procedure continues as follows.

First, predict a new target state estimate $\hat{X}(t_{k+1} \mid t_k)$ of the state $X(t_{k+1})$ at time t_{k+1} given all measurements up to time t_k

$$\hat{X}(t_{k+1} \mid t_k) = \phi(t_k)X(t_k) + A_p(t_k) \tag{7.29}$$

along with its covariance

$$P(k+1 \mid k) = \phi(t_k)P(k \mid k)\phi(t_k)^T + Q(t_k) \tag{7.30}$$

Then, update the target state using the $(k + 1)$st radar measurement

$$\hat{X}(t_{k+1} \mid t_{k+1}) = \hat{X}(t_{k+1} \mid t_k) + K_k[Y_{k+1} - H(t_{k+1})\hat{X}(t_{k+1} \mid t_k)] \tag{7.31}$$

and its covariance

$$P(k+1 \mid k+1) = [I - K_{k+1}H(t_{k+1})]P(k+1 \mid k) \tag{7.32}$$

using the Kalman gains

$$K_{k+1} = P(k+1\,|\,k)H^T(t_{k+1})[H(t_{k+1})P(k+1\,|\,k)H^T(t_{k+1}) + \Re_k]^{-1} \tag{7.33}$$

Because the gains are calculated using the history of all past update times and accuracies, the gains automatically increase after missed detections and automatically increase to give greater weight to a detection when it is known to be more accurate, and they automatically decrease as the track ages, reflecting the value of the detections already filtered. For example, for a zero random acceleration, $Q_k = 0$, and a constant detection covariance matrix, $\Re_k$, the $\alpha-\beta$ filter can be made equivalent to the Kalman filter by setting

$$\alpha = \frac{2(2k-1)}{k(k+1)} \tag{7.34}$$

and

$$\beta = \frac{6}{k(k+1)} \tag{7.35}$$

on the kth scan. Thus, as time passes, α and β approach zero, applying heavy filtering to the new samples. In practical radar applications $Q_k > 0$, and so the tracking gains eventually settle to a non-zero value termed the *steady-state tracking gains.*

The tradeoffs for employing a Kalman filter for radar tracking generally are tuning the filter for the desired degree of filtering, selecting the tracking coordinates, and adapting the filter to deal with changes in the target motion (e.g., maneuvers, different phases of ballistic flight, and so on).

Tuning the Kalman Filter. The greatest advantage of the Kalman filter for radar tracking is that it provides a systematic way of calculating gains. However, a disadvantage is that this gain calculation assumes linear target motion with random perturbations (Eq. 7.22). Most practical radar-tracking problems involve targets that deviate from linear motion in more complex ways (e.g., course corrections, terrain following, evasive maneuvers, and atmospheric drag). The Kalman filter is tuned to a practical radar-tracking problem through the selection of the covariance matrix, $Q(t_k)$, of the unknown random maneuver. The goal of this selection is to obtain the best possible tracking performance for the more complex cases of interest while still using the simple Kalman random perturbation model. For example, in the simplified case of a single dimension and constant tracking conditions, the measurement covariance matrix is simply a single, constant measurement variance, $\Re_k = \sigma_m^2$, and the time between detections is a constant $R_k = T$. In this case, the Kalman filter described in Eqs. 7.29 to 7.33 has gains that are a function of the dimensionless track-filtering parameter γ_{track}:

$$\gamma_{\text{track}} = \frac{qT^3}{\sigma_m^2} \tag{7.36}$$

Because the radar measurement accuracy, as represented by the covariance matrix $\Re$, and the time between detection opportunities, T, are parameters of the radar design itself, the selection of $Q(t_k)$ is the degree of freedom available to the tracking filter design. Table 7.4 summarizes the methods for tuning the Kalman filter.

TABLE 7.4 Comparison of Methods of Tuning Kalman Filter for Practical Radar Tracking Problems

Maneuver Model	Q-submatrix	Steady-state Gain Relation and Tracking Index	Tuning Method	Characteristics
Model no. 1: White noise (spectral density q g^2/Hz) acceleration sampled by radar measurement[45].	$q\begin{vmatrix} T_k^3/3 & T_k^2 \\ T_k^2 & T_k \end{vmatrix}$	$\beta = 6-3\alpha-\sqrt{3\alpha^2-36\alpha+36}$ and $\gamma_{\text{track}} = \dfrac{qT^3}{\sigma_m^2}$	Vary q to increase/decrease gains and obtain desired performance using equations in Table 7.3.	Accommodates variable measurement rates well. Responds to maneuvers but not at the edge of filter stability.
Model no. 2: Random change in acceleration at each measurement interval. Standard deviation of acceleration change is σ_a.[46,47]	$\sigma_a^2\begin{vmatrix} T_k^4 & T_k^3/2 \\ T_k^3/2 & T_k^2 \end{vmatrix}$	$\beta = 2(2-\alpha)-4\sqrt{1-\alpha}$ and $\gamma_{\text{track}} = \dfrac{\sigma_a^2 T^4}{\sigma_m^2}$	Vary σ_a to increase/decrease gains and obtain desired performance using equations in Table 7.3.	Responds very well to maneuvers, but operates at the edge of filter stability. Higher radar measurement rate can actually result in less accurate track.[48]
Model no. 3: Random change in velocity at each measurement interval.[49]	$\sigma_v^2\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}$	$\beta = \dfrac{\alpha^2}{(2-\alpha)}$ and $\gamma_{\text{track}} = \dfrac{\sigma_v^2 T^2}{\sigma_m^2}$	Vary σ_v to increase/decrease gains and obtain desired performance using equations in Table 7.3.	Very conservative with respect to filter stability.

TABLE 7.4 Comparison of Methods of Tuning Kalman Filter for Practical Radar Tracking Problems (*Continued*)

Maneuver Model	Q-submatrix	Steady-state Gain Relation and Tracking Index	Tuning Method	Characteristics
Model no. 4: Constantly accelerating target with a white noise jerk j [(g/s)^2/Hz] sampled by radar measurement. (Jerk is the rate of change of acceleration.) [50,51]	$j\begin{vmatrix} \frac{T_k^5}{20} & \frac{T_k^4}{8} & \frac{T_k^3}{6} \\ \frac{T_k^4}{8} & \frac{T_k^3}{3} & \frac{T_k^2}{2} \\ \frac{T_k^3}{6} & \frac{T_k^2}{2} & T_k \end{vmatrix}$	Steady-state gain calculations described in Fitzgerald.[49] and $\gamma_{\text{track}} = \frac{jT^5}{\sigma_m^2}$.	Select this model when target is known/ expected to be accelerating.	Zero lags to constant acceleration; however, noise errors are much greater.[50]
Model no. 5: Constant, deterministic acceleration a (g). Filter objective is to minimize lag plus c standard deviations.[51,52]	Q submatrix not applicable. Instead, assume constant parabolic motion $\frac{1}{2}at^2$.	$\beta = 2(2-\alpha) - 4\sqrt{1-\alpha}$ and $\gamma_{\text{track}} = \frac{a^2T^4}{c^2\sigma_m^2}$.	Vary a to increase/ decrease gains and obtain desired performance using equations in Table 7.3.	Filter minimizes error for a worst-case deterministic maneuver vice a random one.

As seen in Figure 7.28, the selection of $Q(t_k)$ and thus γ_{track} allows one to uniquely determine the steady-state tracking gains as a function of γ_{track}. One can see that large assumed maneuvers (large q, α_a, or a); larger time between updates, T; or very accurate radar measurements (small $\Re$) will result in large tracking gains. The position gain α is nearly identical for the $Q(t_k)$ models no. 1, 2, 3, and 5 in Table 7.4. However, the velocity gain β differs considerably. For random changes in acceleration at each measurement interval (model no. 2), the gains increase to $(\alpha, \beta) = (1,2)$, which is the limit of filter stability. Thus, this model produces filter gains that are the most aggressive at minimizing lags to maneuvers—at the expense of larger

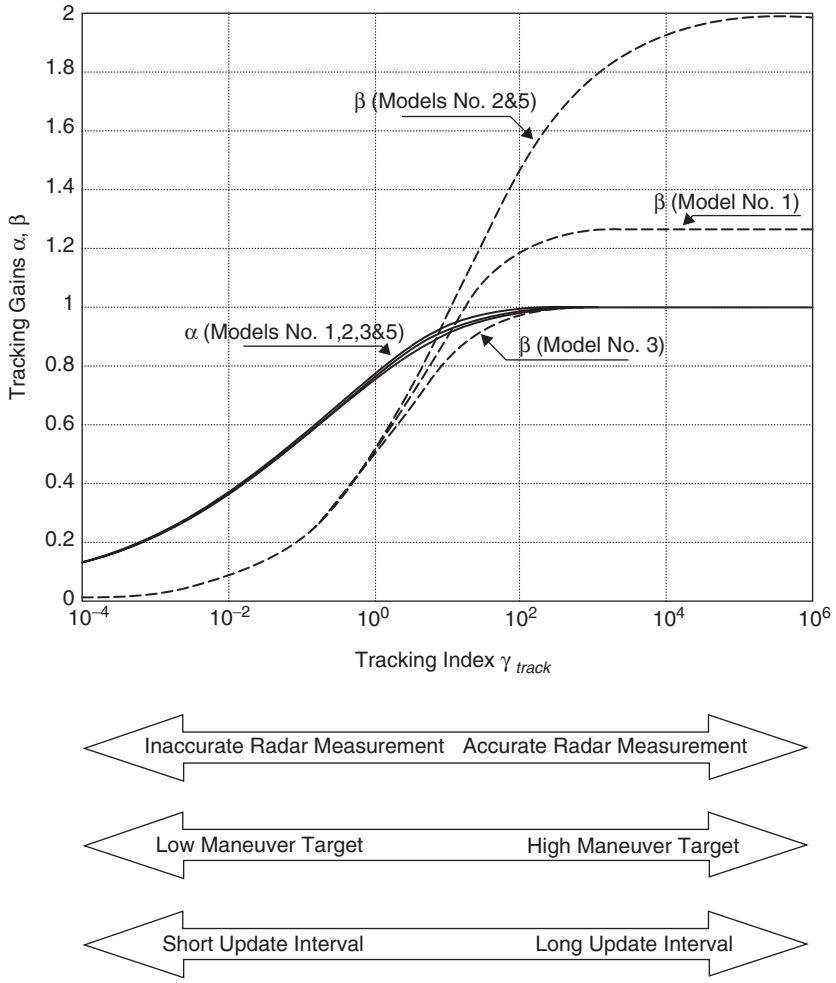


FIGURE 7.28 The relationship between the steady-state tracking gains α and β is shown for different $Q(t_k)$ s corresponding to different assumptions about the unknown target maneuver. Model no. 1: white noise acceleration sampled at each measurement interval; model no. 2: random change in acceleration at each measurement interval; model no. 3: random change in velocity at each measurement interval; and model no. 5: constant deterministic acceleration. Model no. 4 not shown as it is a 3 gain model.

tracking errors due to radar measurement noise. For random changes in velocity at each measurement interval (model no. 3), the gains increase to $(\alpha, \beta) = (1,1)$, which is very conservative from a filter stability point of view. For white noise acceleration sampled by radar measurements (model no. 1), the gains are a compromise, increasing to $(\alpha,\beta) = (1,(3-\sqrt{3}))$. Because this model is a sampled continuous time acceleration, it is preferred when update times are variable because the target does not maneuver more or less when the update interval changes.

The equations in Table 7.3 can then be used to calculate the filter performance in terms of variance reduction ratios and tracking lags. Adjustments to parameters of γ_{track} can be made to obtain the desired noise and lag tradeoff.

Selection of Tracking Coordinates. The Kalman filter assumes linear target motion and a linear relation between the radar detections and the target coordinates. However, radars make detections in polar coordinates (range, angle, doppler), while target motion is most likely linear in Cartesian coordinates (x, y, z). Therefore, some compromises must generally be made in selecting a coordinate system for filtering. Table 7.5 describes the design tradeoffs for different selections.

The polar Kalman filter is rarely used because of the pseudo-accelerations introduced by propagating the state in polar coordinates. The Cartesian/Earth-centered Kalman filter can work well but may have difficulty accommodating radar measurements of less than three dimensions. The extended/dual coordinate system Kalman filter prevents pseudo-accelerations and accommodates measurements of any dimensionality. Both the Cartesian/Earth-centered Kalman filters involve nonlinear transformations resulting in an imperfect calculation of the tracking accuracy. When prediction times are long and/or when very accurate results are needed, these imperfections in the Kalman filter covariance calculation can be significant, and the tracking errors can be quite non-gaussian. Particle filters typically propagate a large number of random samples (particles) from a state transition prior distribution to estimate posterior distributions that are not required to be gaussian in form. Thus, in a particle filter, even multi-modal distributions can be used as prior and realized as posterior distributions. However, particle filters require quite a bit of computation.

The unscented Kalman filter[55] more efficiently calculates the tracking accuracy by propagating selected cardinal points through the filter. The unscented Kalman Filter approximates the covariance matrix with a set of $2L + 1$ sample points, where L is the number of state dimensions. The sample points are propagated through an arbitrary transform function and then used to reconstruct a gaussian covariance matrix. This technique has the advantage of representing the covariance accurately to the third order of a Taylor series expansion. As a result, the calculated tracking accuracy is (at least to third order) uncontaminated (or "unscented") by the nonlinearity.

Adapting Filter to Deal with Changes in Target Motion. The Kalman filter assumes linear target motion perturbed by a random maneuver model as a mathematical convenience in calculating tracking gains. However, most radar targets do not move in a random maneuver but instead move linearly at times and then maneuver unpredictably at times. The challenge in adapting the filter to deal with changes in the target motion (e.g., maneuvers, ballistic re-entry) is to adapt the target motion model for the Kalman filter over time so that more accurate tracking occurs than with a single model. The simplest form of adaptation is a maneuver detector to monitor the tracking filter residuals (differences between measured and predicted position). Large, correlated

TABLE 7.5 Advantages and Disadvantages of Employing the Kalman Filter in Different Coordinate Frames

Kalman Filter Coordinate Frame Variants	Coordinates for Gain Calculation (Eqs. 7.32, 7.33) and state update (Eq. 7.31)	Coordinates for State Prediction (Eqs. 7.29, 7.30)	Method of Covariance Propagation	Advantages	Disadvantages
Polar Kalman filter	Polar	Polar	Eqs. 7.29 to 7.33 in polar coordinates	Filter covariances are calculated exactly and state errors gaussian distributed. Radar detections of less than three dimensions can be used.	Pseudo-accelerations introduced in state propagation.
Cartesian/ Earth-Centered Kalman filter[53]	Cartesian/ Earth-centered	Cartesian/ Earth-centered	Eqs. 7.29 to 7.33 in Cartesian/ Earth-centered coordinates	State propagation is linear (no pseudo-accelerations).	Filter covariances are not exact due to nonlinear transformation.
Extended/ dual coordinate Kalman filter[54]	Polar	Cartesian/ Earth-centered	Eqs. 7.29 to 7.33 in polar coordinates	State propagation is linear (no pseudo-accelerations). Radar detections of less than three dimensions can be easily accommodated.	Requires frequent coordinate transforms.
Unscented Kalman filter[55]	Polar or Cartesian/ Earth-centered	Cartesian/ Earth-centered	Covariance inferred by propagating multiple states	State propagation is linear (no pseudo-accelerations). Filter covariance more exact than traditional methods—particularly for long extrapolation times.	More complex but not necessarily more computation.

residuals generally indicate a maneuver (a deviation from the filter model). Upon maneuver detection, the maneuver spectral density, q, is increased in the Kalman filter model, resulting in higher tracking gains and better following of the maneuver.

A more complex approach is to use multiple Kalman filters running simultaneously with different target motion models—generally, different q values or different equations for target motion (e.g., constant acceleration or constant velocity). Figure 7.29 shows a bank of multiple parallel filters all fed by the same stream of associated measurements. At each detection time, t_k, one of the several filter outputs must be selected to be the track state used for detection to track association.

A systematic way of employing multiple target motion models is the Interacting Multiple Model (IMM) approach diagrammed in Figure 7.30.[56] Multiple models run simultaneously; however, they do not run independently. Instead, there is mixing of

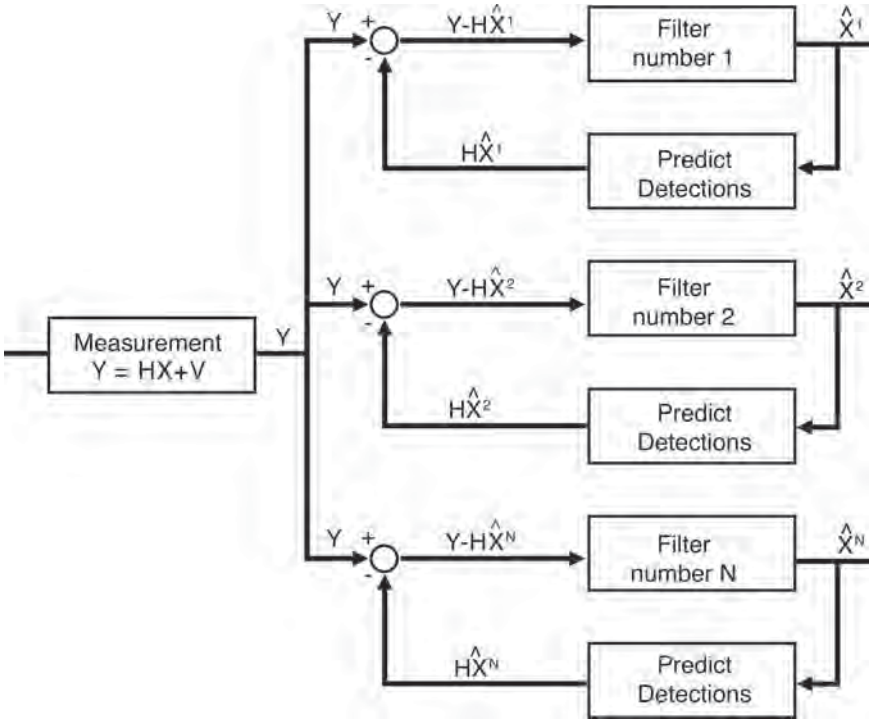


FIGURE 7.29 Bank of parallel radar-tracking filters, each employing a different target motion model (*after S. Blackman and R. Popoli*[44] *© Artech House 1999*)

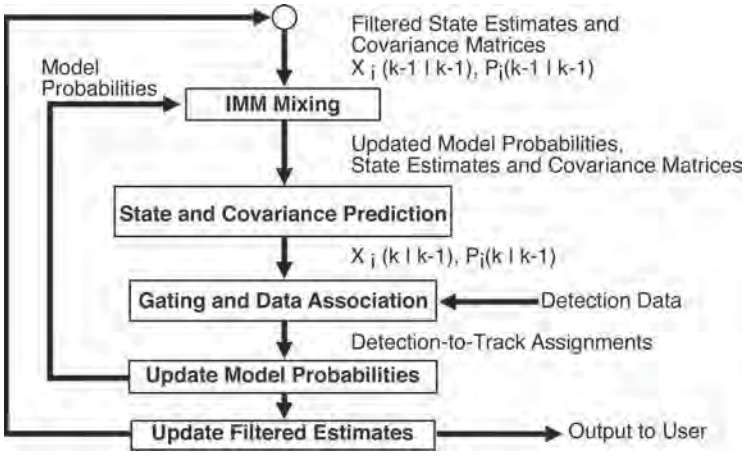


FIGURE 7.30 Flowchart of interacting multiple models (*after S. Blackman and R. Popoli*[44] *© Artech House 1999*)

the model states. The update equation for the *i*th model depends not only on the ith model state but also on the states of all other models. These states are mixed using inferred probabilities of the target transitioning from one motion model to another.

As an example, consider radar tracking of a ballistic missile that undergoes distinct phases of flight: boost, exo-atmospheric flight, and endo-atmospheric re-entry. Each of these phases of flight has a distinct target model.[57] During boost, the target is continually accelerating and increasing speed. This acceleration is unknown and must be estimated. During exo-atmospheric flight, the object is falling with the known acceleration of gravity. During endo-atmospheric re-entry, the target continues to fall but experiences a drag acceleration due to its ballistic coefficient (an unknown target parameter related to the shape and mass of the target). An IMM filter can be used to systematically transition between these different phases of flight, providing a single filter output. Figure 7.31 shows the model probabilities for such an IMM filter application.

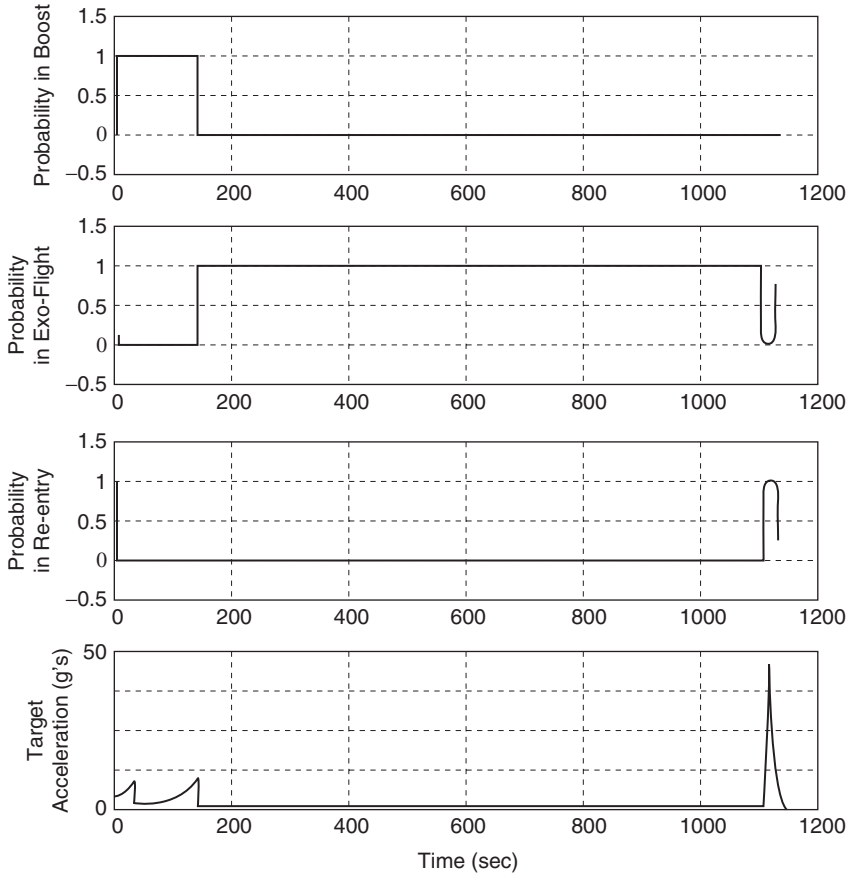


FIGURE 7.31 Model probabilities resulting from the application of an IMM filter to a ballistic missile tracking problem: (*a*) probability that target motion is "boost phase," (*b*) probability that target motion is "exo-atmospheric" flight, (*c*) probability that target motion is "endo-atmospheric" re-entry. (*after R. Coopermanr*[57] *© Fifth International Conference on Information Fusion, vol. 2, 2002*)

Association of Accepted Detection with Existing Tracks. The goal of detection-to-track association is to correctly assign radar detections to existing tracks so the track states in the track file can be correctly updated. The basis for assignment is a measure of how close together the detection and track are in terms of measurable parameters such as range, angle, doppler, and, when available, target signature. The statistical distance is calculated as a weighted combination of the available detection-to-track coordinate differences. In the most general case, this is a complex quadratic form:

$$D^2 = (Y_{k+1} - h(\hat{X}(t_{k+1} \mid t_k))[H(t_{k+1})P(k+1 \mid k)H^T(t_{k+1}) + R_k]^{-1}(Y_{k+1} - h(\hat{X}(t_{k+1} \mid t_k))^T \tag{7.37}$$

For most single radar-tracking problems, it reduces to a simple weighted sum

$$D^2 = \frac{(r_m - r_p)^2}{\sigma_r^2 + \sigma_{\text{pr}}^2} + \frac{(\theta_m - \theta_p)^2}{\sigma_\theta^2 + \sigma_{p\theta}^2} + \frac{(\varphi_m - \varphi_p)^2}{\sigma_\varphi^2 + \sigma_{p\varphi}^2} + \frac{(D_m - D_p)^2}{\sigma_D^2 + \sigma_{\text{pD}}^2} \tag{7.38}$$

where (r_m, θ_m, φ_m, D_m) are the measured range, azimuth, elevation, and doppler with accuracies (σ_r, σ_θ, σ_φ, σ_D); (r_p, θ_p, φ_p, D_p) are the range, azimuth, elevation, and doppler predicted by the automatic tracker with accuracies (σ_{pr}, $\sigma_{p\theta}$, $\sigma_{p\varphi}$, σ_{pD}). The predicted accuracies are a byproduct of the radar tracking filter. Statistical distance rather than Euclidean distance must be used because the range accuracy is usually much better than the azimuth accuracy.

When targets are widely spaced and in a clear environment, only one target detection pair has a small D^2, making these assignments obvious. Thus, the design of detection-to-track association is usually dominated by the more difficult conditions of closely spaced targets or closely spaced targets and clutter. Figure 7.32 shows a common situation for closely spaced targets and/or clutter. Three association gates are constructed around the predicted positions of three existing tracks. Three detections are made, but assignment of the detections to the tracks is not obvious: two detections are within gate 1; three detections are within gate 2; and one detection is within gate 3. Table 7.6 lists all detections

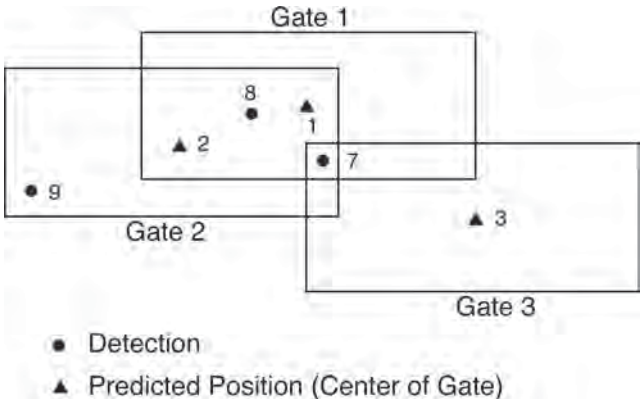


FIGURE 7.32 Examples of the problems caused by multiple detections and tracks in close vicinity (*from G. V. Trunk*[7])

TABLE 7.6 Association Table for Example Shown in Figure 7.32*

Track No.	Detection No. 7	Detection No. 8	Detection No. 9
1	4.2	1.2	∞
2	5.4	3.1	7.2
3	6.3	∞	∞

*(*from G. V. Trunk*[7])

within the tracking gates and the statistical distance between the detection and track. If the detection is outside the track gate, the statistical distance is set to infinity.

Nearest-neighbor assignment is the most common solution to this problem. The simplest form of nearest neighbor works sequentially on incoming data. As each new detection is made, it is assigned to the track with which it has the smallest statistical distance. Hence, if detection no. 9 was received first, it would be assigned to track no. 2. However, it is better to delay the association process slightly so that all detections in a local neighborhood are received and stored and an association table, such as Table 7.6, generated. (This has implications about how sectors are scanned with a phased array.)

Nearest-neighbor assignment can now be applied to the association table by finding the smallest statistical distance between a detection and a track, making that association, and eliminating that detection and track (row and column) from the table. This process is repeated until there are either no tracks or no detections left. Applying this algorithm to Table 7.6 results in detection no. 8 updating track no. 1, detection no. 7 updating track no. 2, and track no. 3 not being updated. Better assignments are possible with more sophisticated processing algorithms. The three types of more sophisticated algorithms most frequently used are

1. *Global Nearest Neighbor (GNN)* Consider the whole matrix of statistical distances simultaneously, and minimize a metric such as the sum of all statistical distances for a complete assignment solution. Performing this optimization can be done using Munkres algorithm.[58] Munkres algorithm is an exact solution of the minimization problem but is rarely used because it is computationally slow. A more computationally efficient exact solution is the Jonker, Volgenant, Castanon (JVC) algorithm.[59] The JVC is much more efficient for sparse assignment matrices (which are likely for practical radar-tracking problems). Speed improvements of 30 to 1000 times have been reported. An effective suboptimal solution is the Auction algorithm, which views the tracks as being "auctioned off" to the detections—iteratively assigning higher costs to tracks competed for by more detections.[60] Figure 7.33 provides a comparison[61] of the Munkres, JVC, and Auction algorithms optimized for sparse data. The JVC and Auction algorithms provide a significant increase in computational speed. Although the Auction algorithm is simpler, requiring less lines of code, the JVC algorithm generally requires less computation time.
2. *Probabilistic Data Association (PDA)* Another alternative is the probabilistic data association (PDA) algorithm[62,63,64] where no attempt is made to assign tracks to detections, but instead, tracks are updated with all the nearby detections—weighted by the perceived probability of the track being the correct association. Because PDA relies on erroneous associations essentially "averaging out," it is most effective when tracks are far enough apart that nearby detections originate from spatially random noise or clutter exclusively and when the tracking gains are small (i.e., when the tracking index γ_{track} is small). The Joint Probabilistic Data

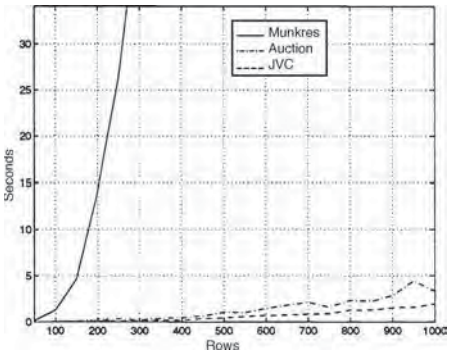


FIGURE 7.33 A comparison of the execution time for the Munkres (optimum), JVC (optimum), and Auction (suboptimum) algorithms shows the rapid increase in computation required for Munkres as the number of rows in the assignment matrix increases. The JVC and auction algorithms show much more gradual growth. (*after I Kadar et al.*[61] *© SPIE.1997*)

Association (JPDA)[65] is an extension of PDA that handles more closely spaced targets. In JPDA, detections are weighted less when they are near another track.

3. *Multiple Hypothesis Algorithms* The most sophisticated algorithms are multiple hypothesis algorithms in which all (or many) possible tracks are formed and updated with each possible detection.[66,67,68] In Table 7.6, track no. 1 would become three tracks (or hypotheses), corresponding to updating with detection no. 8, detection no. 7, and no detection. Each of these tracks would undergo a Kalman filter update and be eligible for association with the next set of detections. Tracks are pruned away in a systematic manner leaving only the most probable. Figure 7.34 illustrates the tracking of a single target using multiple hypothesis techniques. In this example, many hypotheses are formed and, over successive measurement intervals, successfully pruned away leaving only one correct track.

The region of applicability for the more sophisticated algorithms is determined by two parameters: the density of extraneous detections λ (detections per unit area or volume)

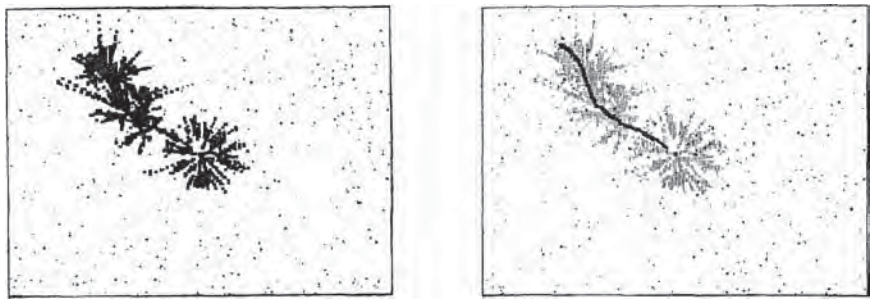


FIGURE 7.34 Example of the use of multiple hypothesis tracking on 90 scans of simulated radar data containing a single target and many false alarms: (*a*) shows all hypotheses forms and (*b*) shows the single hypothesis selected. (Pruned hypotheses are grayed out.) (*after W. Koch*[69] *© IEEE 1995*)

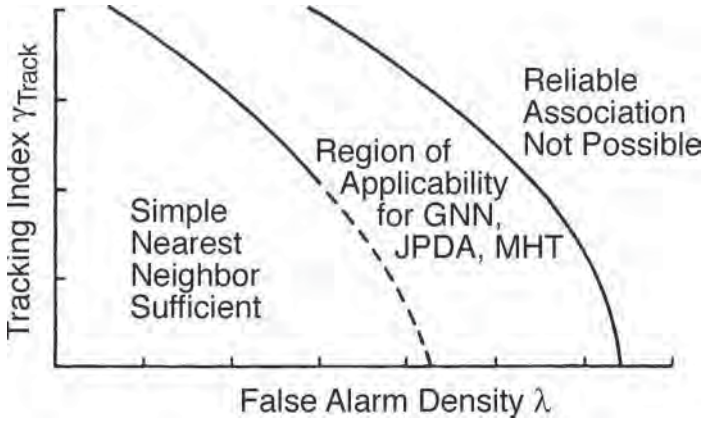


FIGURE 7.35 The applicability of different detection-to-track association algorithms is determined by the density of false alarms and the dimensionless tracking parameter γ_{track}. (*after D. J. Salmond*[70] *© SPIE 1990*)

and the dimensionless track filtering parameter γ_{track}. Figure 7.35 bounds this region of applicability. When λ and γ_{track} are small, then there is no need for any more than simple nearest neighbor tracking, and indeed, most tracking systems still use this approach. As λ increases, there is greater risk of false association decisions; however, the effect of this is reduced if γ_{track} is small. At the other extreme, when λ and γ_{track} are large, the tracking problem is essentially unsolvable without basic changes to the radar design parameters to reduce them. There is an intermediate region where sophisticated association has value. The width of this region is very specific to the particular problem. When γ_{track} is large and very little delay in the output can be tolerated, then the region of applicability is fairly small and very simple multiple hypothesis approaches (splitting tracks into at most one or two hypotheses) are then the best answer.

When γ_{track} is small, then PDA/JPDA can be used to operate at significantly higher false alarm densities. When significant delay can be tolerated in the output, then many hypotheses can be formed (as in Figure 7.34) and orders of magnitude more detections handled. Blackman and Popoli[44] provide a good survey of comparative studies in this area. One study using data recorded from flights of closely spaced aircraft showed very little difference between GNN, JPDA, and MHT.[38] However, theoretical predictions can show differences of orders of magnitude in the density of clutter detections that can be handled.[44]

New Track Formation. There are two classes of track formation algorithms:

1. *Forward-tracking algorithms* basically propagate one hypothesis forward in time, recursively checking for "target-like" motion. Detections that do not correlate with clutter points or tracks are used to initiate new tracks. If the detection does not contain doppler information, the new detection is usually used as the predicted position (in some military systems, one assumes a radially inbound velocity), and a large correlation region must be used for the next observation. The correlation region must be large enough to capture the next detection of the target, assuming that it could have the maximum velocity of interest. A common track initiation

criterion is four out of five, although one may require only three detections out of five opportunities in regions with a low false-alarm rate and a low target density. However, one may require a much larger number of detections when the radar has the flexibility of an electronic scan that can place many detection opportunities in a short time interval.

2. *Backward-tracking or "batch" algorithms* consider all the detections simultaneously, attempting to match the detections to a "target-like" pattern. This can be done by actually constructing a large number of matched filters, as in retrospective processing[71] (see Figure 7.36), or by using a forward-tracking process with multiple hypothesis formed and propagated.

Just as automatic radar detection is a tradeoff between probability of detection and probability of false alarm, new track formation is a tradeoff between the speed at which a track is formed and the probability of erroneously forming a false track that does not represent a physical object of interest. There are two types of false tracks: (1) Tracks on real objects that are simply not of interest. For example, if the targets of interest are airplanes, then a false track could be a track on a bird. (2) Tracks composed of unrelated detections from different objects that the automatic tracking process has mistakenly associated together. For example, a false track could be composed of detections from several different stationary clutter points that have been associated together over time to create a false moving track.

The approach for preventing false tracks on objects not of interest is to actually develop tracks on all of them but then observe them long enough to classify them as unwanted. In the case of the bird, one would gather enough detections to improve the velocity accuracy of the track so that it is clear whether the track is of interest or not. Thus, one desires to delay the disclosure of a track until enough time has passed to classify it accurately. This accuracy can be determined by T_{obs}, the amount of time over which the object is observed and by basic parameters of the radar:

T = the time between successive detections

σ = the accuracy in a particular dimension of interest

M = the number of detections used in forming the track

$N \cong (T_{\text{obs}}/T) + 1$, which is the number of detection opportunities

The velocity accuracy is given by the following equation:[72]

$$\sigma_v = \frac{\sigma}{T_{\text{obs}}} \times \left[\frac{12(N-1)}{N(N+1)}\right]^{1/2} \tag{7.39}$$

The dominant design parameters in the equation are the accuracy of the radar and the observation time. (Better accuracy or longer observation time allows more accurate measurement of velocity.) Making more detections in the observation time improves the accuracy but only in a square-root sense.

The approach to preventing falsely composed tracks from different objects in a clutter region, G, is to require enough detections in a tight enough pattern to make $E[N_{\text{FT}}]$, the expected number of false tracks, small. When there is an average of N_C detections in a D-dimensional region G, then[73]

$$E[N_{\text{FT}}] = \lambda_F \times \lambda_p^{M-2} \times N_C^M \times \gamma(D, N, M) \tag{7.40}$$

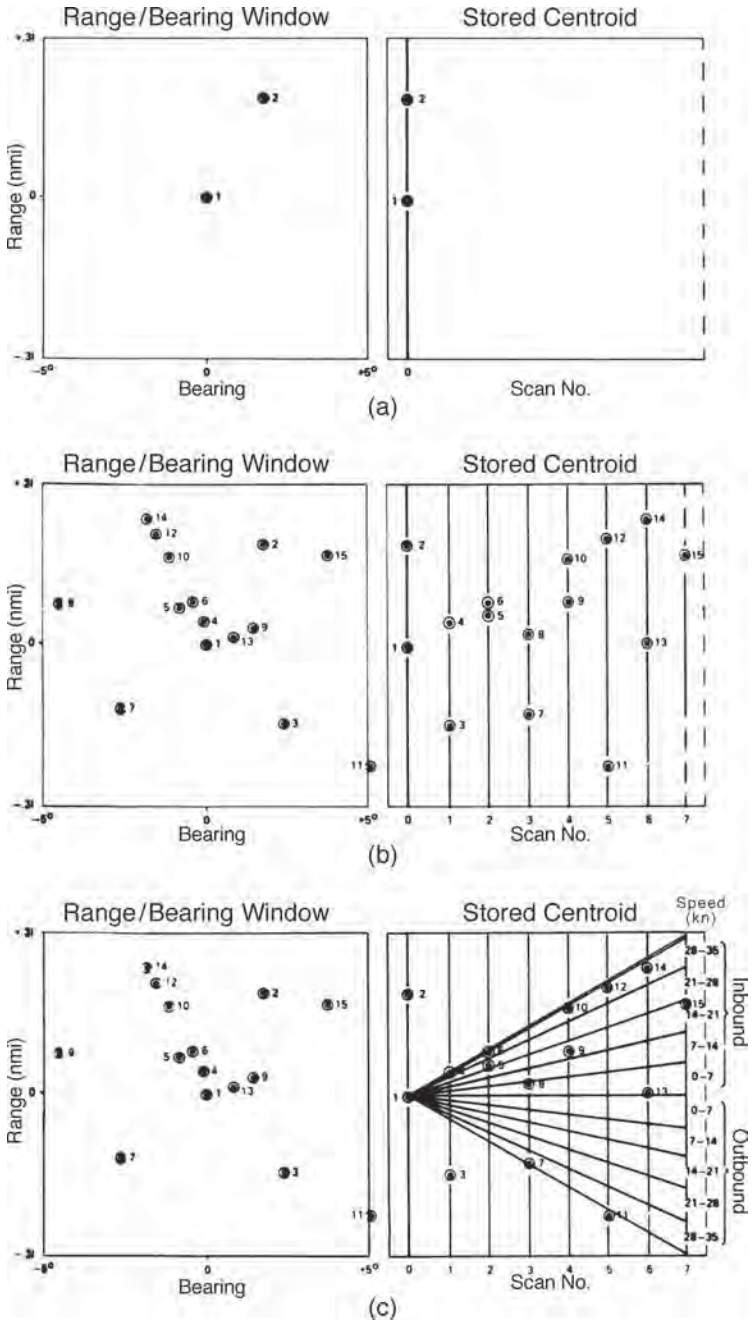


FIGURE 7.36 The retrospective process: (*a*) a single scan of data, (*b*) eight scans of data, and (*c*) eight scans of data with trajectory filters applied (*after Prengaman et al.*[71] *© IEEE 1982*)

where λ_F is the ratio of the size of the possible space a target can travel in one detection interval to the size of entire clutter region *G,*

$$\lambda_F = \frac{(V_{\text{MAX}})^D}{G} \tag{7.41}$$

and λ_P is the ratio of the size of a radar resolution cell to the size of the entire clutter region *G*

$$\lambda_P = \frac{\tau_1 \cdot \cdot \cdot \tau_D}{G} \tag{7.42}$$

τ_i being the resolution "distance" in the *i*th dimension, and $\gamma(D, N, M)$ being the combinatorial term:

$$\gamma(D,N,M) = (N-1)^D \binom{N-1}{M-1} 2^{D(M-1)} \tag{7.43}$$

Figure 7.37 gives an example of the application of Eqs. 7.40 to 7.43 to a radar with $\lambda_P = 2.10^{-3}$ and $\lambda_P = 10^{-5}$. Increasing the number of detections required to form

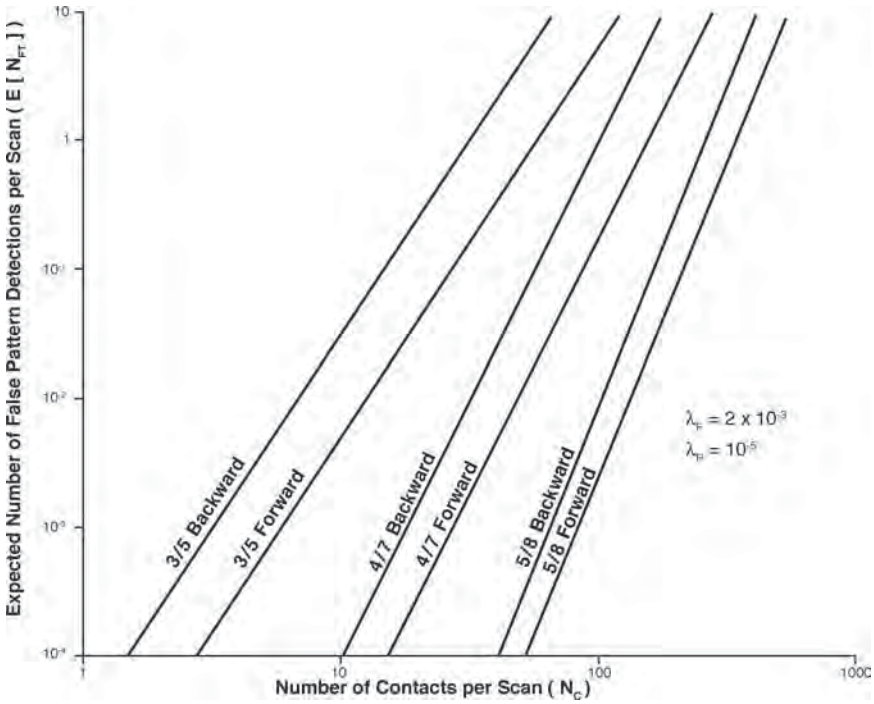


FIGURE 7.37 Variation of the expected number of false tracks with the track formation *M*-out-of-*N* criterion (*after W. G. Bath et al.*[73])

a track from three (out of five) to five (out of eight) increases the density of false alarms that can be tolerated by more than an order of magnitude. Forward- and backward-tracking algorithms produce similar numbers of false tracks. However, the backward-tracking algorithms can operate in more ambiguous situations (where the density of false alarms, λ, is comparable to or greater than λ_F or λ_P). Under these ambiguous circumstances, the forward tracker will have multiple detections in a track formation or promotion gate and will require multiple hypothesis to reliably form tracks.

The design of the track formation process and the automatic detection process should be considered together. A longer time allowed for track formation (higher *M*/*N*) allows the radar detection process to use lower detection thresholds, resulting in better radar sensitivity. For any given set of radar parameters, *M*/*N* track formation criterion, and probability distribution of clutter amplitudes, there exists an optimum false-alarm rate that minimizes the signal-to-noise ratio required to detect targets. Figure 7.38 illustrates this optimization for an eight-scan track formation process.

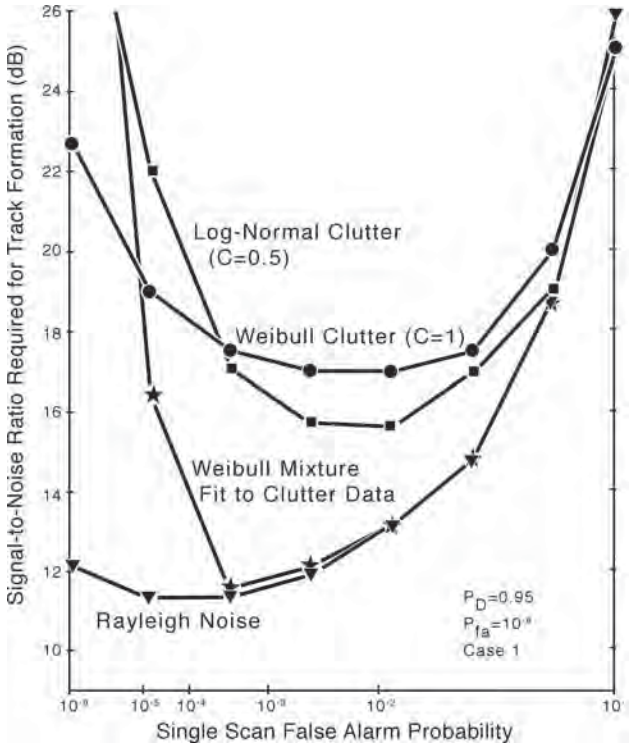


FIGURE 7.38 Overall sensitivity of an automatic detection and automatic tracking process working together. The single-scan false-alarm probability can be optimized to provide the lowest required signal-to-noise ratio for various probability distributions of clutter amplitude. (*after Prengaman et al.*[74] © *IEEE 1982*)

Very low single-scan false-alarm probabilities allow tracks to be formed quickly. However, if a longer delay is tolerable, then detection thresholds can be lower, resulting in better sensitivity in non-gaussian clutter.

Radar Scheduling and Control. The interaction of the radar-tracking system with the scheduling and control function of the radar is minor for mechanical rotating radars but major for phased array radars. For mechanically rotating radars, all that is usually done is that the tracking gates are fed back to the signal processor. The tracking gates are always used to facilitate the association process and may be used to lower the detection threshold within the gate and/or modify the contact entry logic within the gate (e.g., modify rules governing clutter maps).

The interaction of the tracking system with a phased array radar is much more significant. The major benefit of a phased array with respect to tracking is in the area of track initiation.[75,76] Phased arrays use a confirmation strategy to initiate tracks rapidly. That is, after the association process, all unassociated detections generate confirmation dwells to confirm the existence of a new track. The initial confirmation dwell uses the same waveform (frequency and PRF, if a pulse-doppler waveform) but may increase the energy. Analysis has shown that a 3-dB increase in the transmitted confirmation energy (additional energy is also available by placing the target in the center of the confirmation beam) can significantly increase the probability of confirmation.[76] Furthermore, the confirmation dwell should be transmitted as soon as possible to maintain a Swerling I fluctuation model. (That is, if the target was originally detected when the target fluctuation produced a large return, the confirmation dwell will see this same large return.) After confirmation, a series of initial track maintenance dwells over several seconds is used to develop an accurate state vector. A complete discussion of priority associated with tracking within the scheduler of a phased array is beyond the scope of this brief discussion. However, it is worthwhile noting these general rules: (1) Confirmation dwells should have a priority higher than all other functions except those associated with weapon control; (2) low priority tracks (e.g., tracks at long range) can be updated using search detections; and (3) high priority tracks should have a priority higher than volume surveillance. The update rate for high priority tracks should be such that a single tracking dwell is sufficient to update the track. The actual update rate will depend on many factors including (a) maximum target speed and maneuver capability, (b) radar beamwidth (beam could be spoiled), (c) range of the radar track, and (d) accuracy of predicted position. If a pulse doppler dwell is required to update the track in clutter, the waveform should be selected to place the target near the center of the ambiguous range-doppler detection space. Finally, the track can be updated with the ambiguous range-doppler detection because the track state-vector can be used to remove the ambiguity.

7.4 NETWORKED RADARS

Ideally, a single radar can reliably detect and track all targets of interest. However, the environment and the laws of physics often will not permit this. In general, no single radar can provide a complete surveillance and tracking picture. Radar networking can be a good solution to this problem and, in some cases, may be more cost effective than solving the problem through one very high performance radar. Radar networking systems are generally characterized by what radar data are shared and how they are correlated and fused. The two most common ways of combining radar data are as follows:

1. Detection-to-track fusion (see Figure 7.39, upper half) associates each detection to the networked track, calculated potentially using detections from all radars. Thus, the entire stream of detections (up to the present) is potentially available to calculate the track state used for the association decision on the most recent detection.
2. Track-to-track fusion (see Figure 7.39, lower half) associates each detection to a single radar track state calculated using only detections from that radar. The single radar track states are then grouped with each other to produce a netted track state.

The design decision as to which approach is better for grouping data depends on the radars and targets involved. One case where detection-to-track association is clearly better is when the radars have a reduced probability of detection so there are potential gaps in the data stream or periods where the data stream is sparse. In these cases, a much more accurate track state can be calculated using multiple data streams than using only one because multiple streams will tend to fill in the gaps in detection and restore a high consistent data rate during periods of reduced probability of detection. Figure 7.40 illustrates the sensitivity to target fades by plotting the track region of uncertainty (ROU) versus the probability of detection for single radar tracking and multiple radar tracking. The ROU is defined as the distance that contains the error with 99 percent probability and is

$$\text{ROU} = 2.3(\textit{tracking error due to detection noise}) + (\textit{tracking error due to maneuver})$$

This can be calculated for any case of interest using the formulas in Table 7.3.

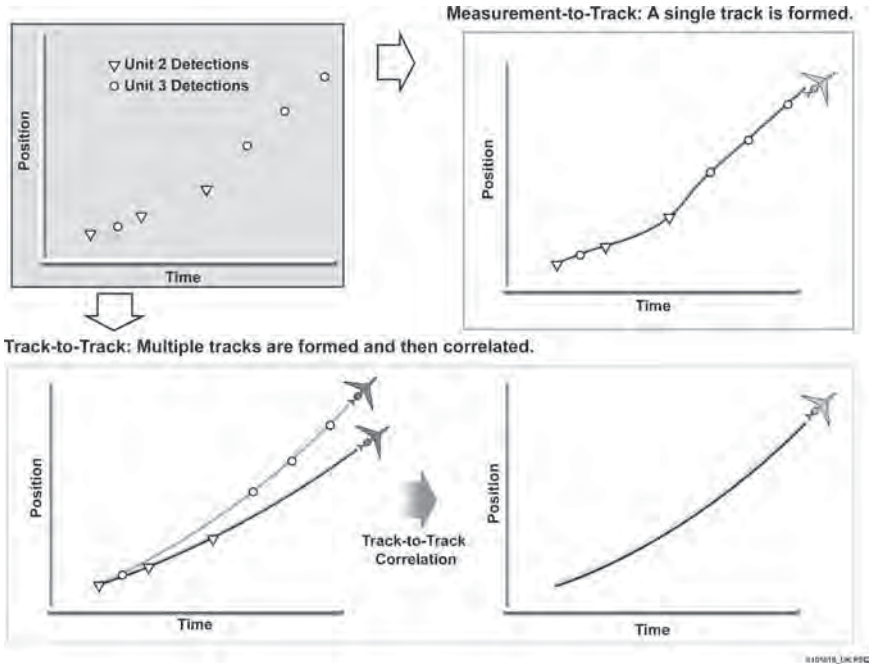


FIGURE 7.39 There are two common methods of fusion data in radar networking: detection-to-track and track-to-track. (*after W. Bath*[77] *© IEE 2002*)

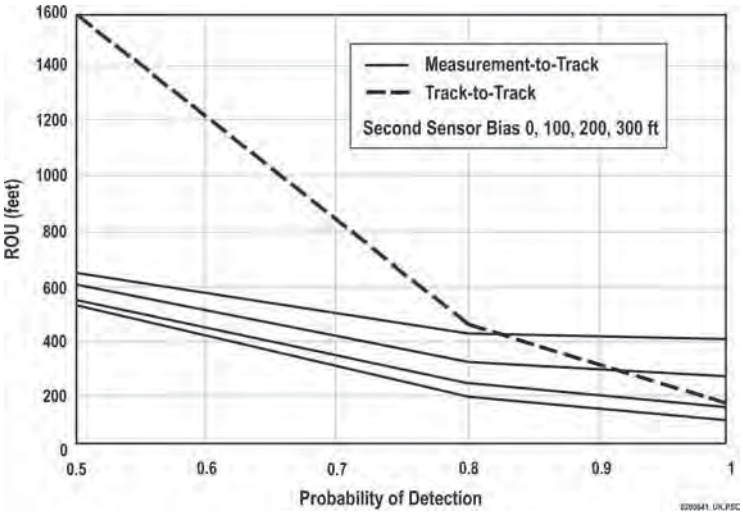


FIGURE 7.40 Comparison of detection-to-track and track-to-track association. For fading targets ($Pd < 1$), detection-to-track is preferred. For large sensor biases and non-fading targets, track-to-track is preferred. (*after W.Bath[77] © IEE 2002*)

When the probability of detection is much less than unity, the measurement-to-track fusion is considerably more accurate. This is easily explained by the fact that the probability of a significant outage of data is much reduced if two sources are available. With a more accurate track, tighter association criteria can be used for detections.

If the biases cannot be effectively removed, then there may be an advantage to associating to a single radar track—which by definition is unbiased with respect to itself. If biases cannot be kept smaller than the ROU, then at high probabilities of detection, one prefers single radar association followed by track-to-track association.

It is possible to make simple comparisons between the accuracy of detection fusion as opposed to track fusion for equivalent use of data bandwidth to exchange radar data. When ROU is plotted as a function of the position gain α, it has the "bathtub" shape shown by the single radar curve in Figure 7.41. The left-hand side of the "bathtub" is dominated by the lag component, while the right-hand side is dominated by the radar measurement noise component. Because the gains (horizontal axis) are the designer's choice, the single radar ROU is the minimum of the "bathtub" curve.

Now consider the fusion of two radars in a particular dimension. If one radar has one-tenth the ROU of the other in this dimension, then the more accurate radar in this dimension will dominate and essentially determine the result. At least in steady state, it is relatively easy to produce this dominance by any of the fusion methods. Of more interest is the case where the radars are comparable in terms of accuracy and update rate, producing comparable ROUs. This case more clearly shows the difference in the fusion methods.

For example, when two identical radars are combined by detection fusion, then the update rate is essentially doubled. This reduces the lag by a factor of 4, allowing a smaller gain to be selected (optimization more to the left of the "bathtub"), reducing the tracking errors due to measurement noise. The net result is the movement from the single radar curve to the detection fusion curve in Figure 7.41.

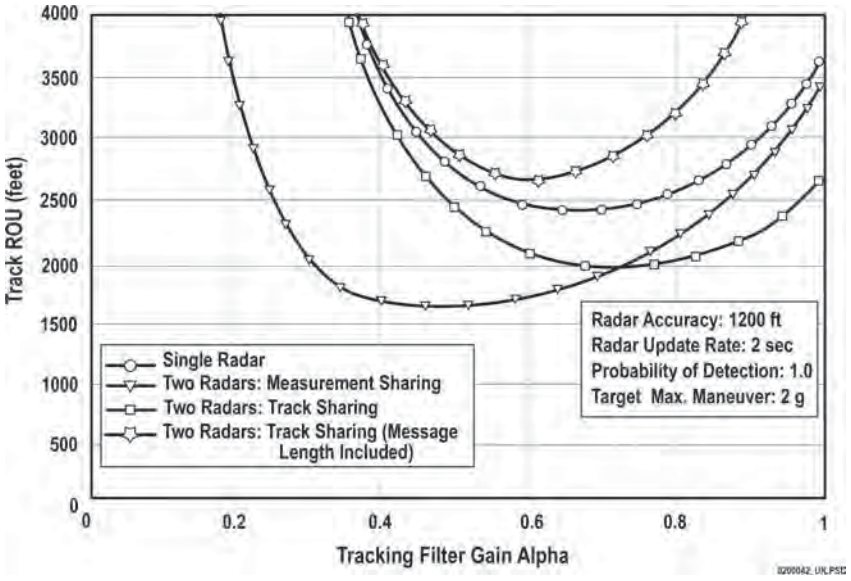


FIGURE 7.41 Comparison of detection fusion and track fusion approaches. For air-breathing targets, detection fusion produces the most accurate track (smallest ROU). (*after W.Bath*[77] *© IEE 2002*)

When two identical radars are combined by track fusion, the update rate for each tracking process does not change, and so the lag does not change. However, the standard deviation of the tracking errors due to measurement noise is reduced by the square root of 2, allowing a larger gain to be selected (optimization more to the right of the bathtub), reducing the lag. The net result is the movement from the single radar curve to the track fusion curve in Figure 7.41.

If there is any significant maneuver possible, the factor of 4 in lag will have a more significant effect than the factor of the square root of 2 in the square root of the tracking errors due to measurement noise. Thus, one can see the detection fusion curve achieves a significantly lower minimum than the track fusion curve.

To combine data from multiple radars, the data must be placed in a common coordinate system. This process is called *grid locking* and involves specifying the location of the radars and estimating radar biases in range and angle. The previous difficult problem of radar location is solved trivially by the global positioning system. An estimate of radar biases between two radars can be obtained from a long-term average of the difference between predicted and measured coordinates on all tracks that have a substantial number of detections from both radars.[78]

7.5 UNLIKE-SENSOR INTEGRATION

A number of sensors can be integrated: radar, identification friend or foe (IFF), the air traffic control radar beacon system (ATCRBS), infrared, optical, and acoustic. The sensors that are most easily integrated are the electromagnetic sensors, i.e., radar, IFF, and strobe extractors of noise sources or emitters.

IFF Integration. The problem of integrating radar and military IFF data is less difficult than that of integrating two radars. The question of whether detections or tracks should be integrated is a function of the application. In a military situation, by integrating detections one could interrogate the target only a few times, identify it, and then associate it with a radar track. From then on, there would be little need for re-interrogating the target. However, in an air traffic control situation using ATCRBS, targets would be interrogated at every scan and, consequently, either detections or tracks could be integrated.

Radar–DF Bearing Strobe Integration. Correlating radar tracks with DF (direction finding) bearing strobes on emitters has been considered by Coleman[79] and later by Trunk and Wilson.[80,81] Trunk and Wilson considered the problem of associating each DF track with either no radar track or one of m radar tracks. In their formulation, there were K DF angle tracks, each specified by a different number of DF detections; and similarly, m radar tracks, each specified by a different number of radar detections. Because each target can carry multiple emitters (i.e., multiple DF tracks can be associated with each radar track), each DF track association can be considered by itself, resulting in K disjoint association problems. Consequently, an equivalent problem is given a DF track specified by n DF bearing detections, one can associate the DF track with no radar track or with one of m radar tracks, the jth radar track being specified by m_j radar detections. Using a combination of Bayes and Neyman-Pearson procedures and assuming that the DF detection errors are usually independent and gaussian-distributed with zero mean and constant variance σ^2 but with occasional outliers (i.e., large errors not described by the gaussian density), Trunk and Wilson argued that the decision should be based on the probability

$$P_j = \text{probability } (Z \geq d_j) \tag{7.44}$$

where Z has a chi-square density with n_j degrees of freedom and d_j is given by

$$d_j = \sum_{i=1}^{n_j} \min\{4,[\theta_e(t_i) - \theta_j(t_i)]^2 / \sigma^2\} \qquad j = 1,\ldots,m \tag{7.45}$$

where n_j is the number of DF detections overlapping the time interval for which the jth radar track exists; $\theta_e(t_i)$ is the DF detection at time t_i; $\theta_j(t_i)$ is the predicted azimuth of radar track j for time t_i; and the factor 4 limits the square error to $4\sigma^2$ to account for DF outliers. By using the two largest P_js, designated Pmax and Pnext, and thresholds T_L, T_H, T_M, and R, the following decisions and decision rules were generated:

1. **Firm correlation** DF signal goes with radar track having largest P_j (i.e., Pmax) when $P\text{max} \geq T_H$ and $P\text{max} \geq P\text{next} + R$.
2. **Tentative correlation** DF signal probably goes with radar track having largest P_j (i.e., Pmax) when $T_H > P\text{max} \geq T_M$ and $P\text{max} \geq P\text{next} + R$.
3. **Tentative correlation with some track** DF signal probably goes with some radar track (but cannot determine which) when $P\text{max} \geq T_M$ but $P\text{max} < P\text{next} + R$.
4. **Tentatively uncorrelated** DF signal probably does not go with any radar track when $T_M > P\text{max} > T_L$.
5. **Firmly uncorrelated** DF signal does not go with any radar track when $T_L \geq P\text{max}$.

The lower threshold T_L determines the probability that the correct radar track (i.e., the one associated with the DF signal) will be incorrectly rejected from further consideration. If the desired rejection rate for the correct track is P_R, one can obtain this by setting $T_L = P_R$. The threshold T_H is set equal to P_{fa}, defined as the probability of falsely associating a radar track with a DF signal when the DF signal does not belong with the radar track. The threshold T_H is a function of the azimuth difference μ between the true (DF) position and the radar track under consideration. The threshold T_H was found for $\mu = 1.0\sigma$ and $\mu = 1.5\sigma$ by simulation techniques, and the results for $P_{\text{fa}} = 0.01$ are shown in Figure 7.42. Between the high and low thresholds, there is a tentative region. The middle threshold divides the "tentative" region into a tentatively correlated region and a tentatively uncorrelated region. The rationale in setting the threshold is to set the two associated error probabilities equal for a particular separation. The threshold T_M was found by using simulation techniques and is also shown in Figure 7.42.

The probability margin R ensures the selection of the proper DF radar association (avoiding rapidly changing decisions) when there are two or more radar tracks close to one another. The correct selection is reached by postponing a decision until the two highest association probabilities differ by R. The value for R is found by specifying a probability of an association error P_e according to $P_e = P\,(P\text{max} \geq P\text{next} + R)$, where Pmax corresponds to an incorrect association and Pnext corresponds to the correct association. The probability margin R is a function of P_e and the separation μ of the radar tracks. The probability margin R was found for $\mu = 0.25\sigma$, 0.50σ, and 1.00σ by using simulation techniques, and the results for $P_e = 0.01$ are shown in Figure 7.43.

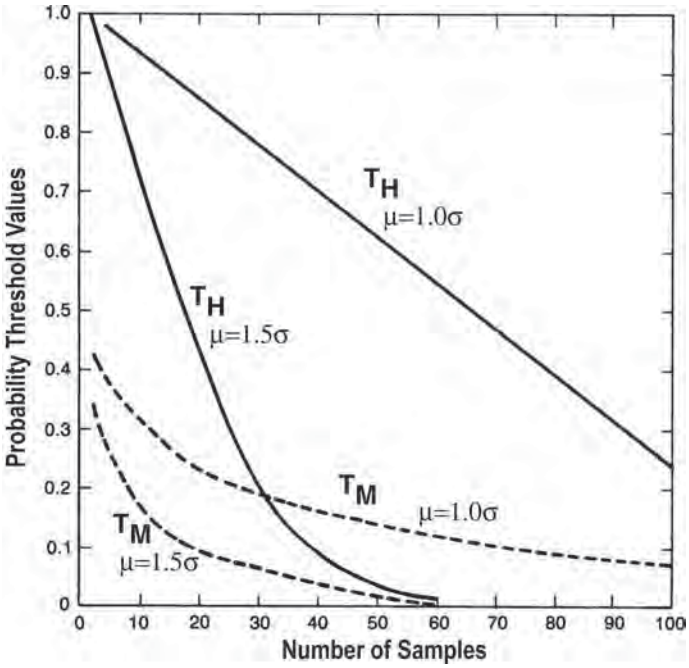


FIGURE 7.42 High threshold (solid lines) and middle threshold (dashed lines) versus number of samples for two different separations (*after G.V. Trunk and J.D. Wilson*[80] *© IEEE 1987*)

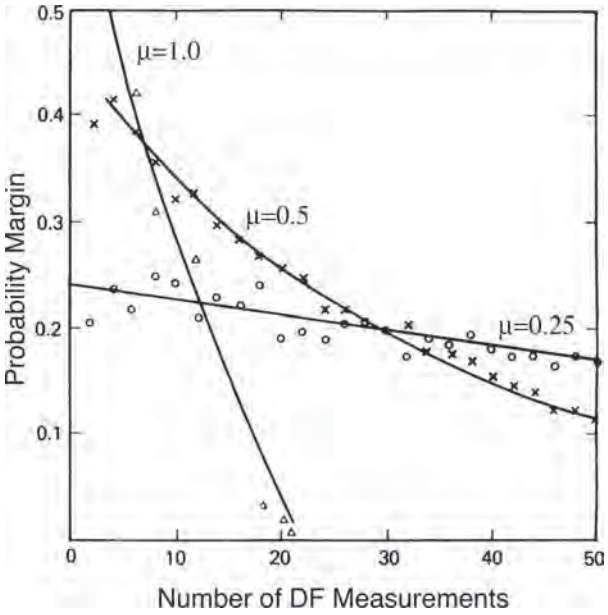


FIGURE 7.43 Probability margin versus number of DF detections for three different target separations. The *o*s, *x*s, and Δs are the simulation results for $\mu = 0.25$, $\mu = 0.5$, and $\mu = 1.0$, respectively. (*after G.V. Trunk and J.D. Wilson*[80] *© IEEE 1987*)

Because the curves cross one another, one can ensure that $P_e \leq 0.01$ for any μ by setting R equal to the maximum value of any curve for each value of n.

The algorithm was evaluated by using simulations and recorded data. When the radar tracks are separated by several standard deviations of the detection error, correct decisions are made rapidly. However, if the radar tracks are close to one another, errors are avoided by postponing the decision until sufficient data are accumulated. An interesting example with recorded data is shown in Figures 7.44 and 7.45. Figure 7.44 shows the radar (azimuth) detections of the control aircraft, the radar detections of four aircraft of opportunity in the vicinity of the control aircraft, and the DF detections on the radar on the control aircraft. The association probabilities, with and without limiting in Eq. 7.45, are shown in Figure 7.45. Initially, an aircraft of opportunity has the highest association probability; however, a firm decision is not made because *P*max does not exceed *P*next by the probability margin. After the 14th DF detection, the emitter is firmly correlated with the control aircraft. However, at the 18th DF detection, a very bad detection (outlier) is made, and the firm correlation is downgraded to a tentative correlation if limiting is not used. If limiting is employed, however, the correct decision remains firm.

In a complex environment where there are many radar tracks and DF signal sources, it is quite possible that many DF signals will be assigned the category that the DF signal probably goes with some radar track. To remove many of these ambiguities, multisite DF operation can be considered. The extension of the previous procedures

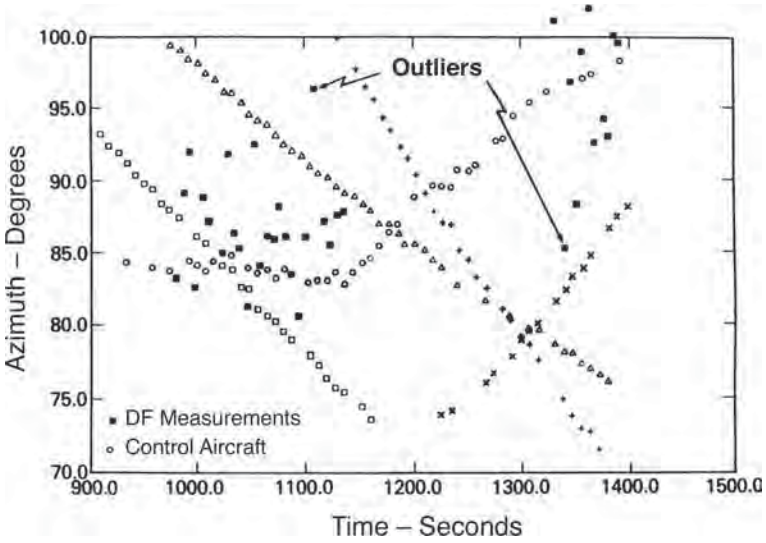


FIGURE 7.44 Radar detections *o* and DF detections collected on the control aircraft. The *o*s, Δs, +s, and *x*s are radar detections on four aircraft of opportunity in the vicinity of the control aircraft. (*after G.V. Trunk and J.D. Wilson*[80] *© IEEE 1987*)

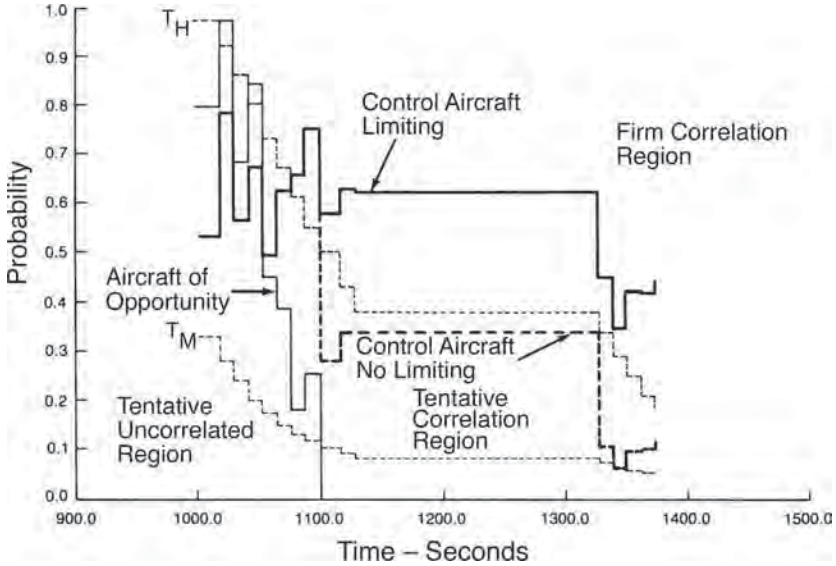


FIGURE 7.45 Association probabilities for experimental data. The bold lines are probabilities for the control aircraft; the solid line, for limiting; the dashed line, for no limiting; the thin line, the association probability for the aircraft of opportunity; and the thin dashed lines, the thresholds T_M and T_H. (*after G.V. Trunk and J.D. Wilson*[80] *© IEEE 1987*)

to multisite operation is straightforward. Specifically, if $\theta_{e1}(t_i)$ and $\theta_{e2}(t_k)$ are the DF angle detections with respect to sites 1 and 2 and if $\theta_{j1}(t_i)$ and $\theta_{j2}(t_k)$ are the estimated angular positions of radar track j with respect to sites 1 and 2, then the multisite squared error is simply

$$d_j = \sum_{i=1}^{n_{1j}} \min\left\{4, [\theta_{e1}(t_i) - \theta_{j1}(t_i)]^2 / \sigma_1{}^2\right\} + \sum_{k=1}^{n_{2j}} \min\left\{4, [\theta_{e2}(t_k) - \theta_{j2}(t_k)]^2 / \sigma_2{}^2\right\} \tag{7.46}$$

The previously described procedure can then be used, with d_j being defined by Eq. 7.46 instead of Eq. 7.45.

REFERENCES

1. J. I. Marcum, "A statistical theory of target detection by pulsed radar," *IRE Trans*., vol. IT-6, pp. 59–267, April 1960.
2. P. Swerling, "Probability of detection for fluctuating targets," *IRE Trans*., vol. IT-6, pp. 269–300, April 1960.
3. J. Neyman and E. S. Pearson, "On the problems of the most efficient tests of statistical hypotheses," *Philos. Trans. R. Soc. London*, vol. 231, ser. A, p. 289, 1933.
4. L. V. Blake, "The effective number of pulses per beamwidth for a scanning radar," *Proc. IRE*, vol. 41, pp. 770–774, June 1953.
5. G. V. Trunk, "Comparison of the collapsing losses in linear and square-law detectors," *Proc. IEEE*, vol. 60, pp. 743–744, June 1972.
6. P. Swerling, "Maximum angular accuracy of a pulsed search radar," *Proc. IRE*, vol. 44, pp. 1146–1155, September 1956.
7. G. V. Trunk, "Survey of radar ADT," Naval Res. Lab. Rept. 8698, June 30, 1983.
8. G. V. Trunk, "Comparison of two scanning radar detectors: The moving window and the feedback integrator," *IEEE Trans*., vol. AES-7, pp. 395–398, March 1971.
9. G. V. Trunk, "Detection results for scanning radars employing feedback integration," *IEEE Trans*., vol. AES-6, pp. 522–527, July 1970.
10. G. V. Trunk and B. H. Cantrell, "Angular accuracy of a scanning radar employing a 2-pole integrator," *IEEE Trans*., vol. AES-9, pp. 649–653, September 1973.
11. B. H. Cantrell and G. V. Trunk, "Corrections to 'angular accuracy of a scanning radar employing a two-pole filter'," *IEEE Trans*., vol. AES-10, pp. 878–880, November 1974.
12. D. C. Cooper and J. W. R. Griffiths, "Video integration in radar and sonar systems," *J. Brit. IRE*, vol. 21, pp. 420–433, May 1961.
13. V. G. Hansen, "Performance of the analog moving window detection," *IEEE Trans*., vol. AES-6, pp. 173–179, March 1970.
14. P. Swerling, "The 'double threshold' method of detection," Project Rand Res. Mem. RM-1008, December 17, 1952.
15. J. V. Harrington, "An analysis of the detection of repeated signals in noise by binary integration," *IRE Trans*., vol. IT-1, pp. 1–9, March 1955.
16. M. Schwartz, "A coincidence procedure for signal detection," *IRE Trans*., vol. It-2, pp. 135–139, December 1956.
17. D. H. Cooper, "Binary quantization of signal amplitudes: effect for radar angular accuracy," *IEEE Trans*., vol. Ane-11, pp. 65–72, March 1964.
18. G. M. Dillard, "A moving-window detector for binary integration," *IEEE Trans*., vol. IT-13, pp. 2–6, January 1967.

19. D. C. Schleher, "Radar detection in log-normal clutter," in *IEEE Int. Radar Conf*., Washington, DC, 1975, pp. 262–267.
20. "Radar processing subsystem evaluation," vol. 1, Johns Hopkins University, Appl. Phys. Lab. Rept. FP8-T-013, November 1975.
21. H. M. Finn and R. S. Johnson, "Adaptive detection mode with threshold control as a function of spacially sampled clutter-level estimates," *RCA Rev*., vol. 29, pp. 141–464, September 1968.
22. R. L. Mitchell and J. F. Walker, "Recursive methods for computing detection probabilities," *IEEE Trans*., vol. AES-7, pp. 671–676, July 1971.
23. G. V. Trunk and J. D. Wilson, "Automatic detector for suppression of sidelobe interference," in *IEEE Conf. Decision & Control*, December 7–9, 1977, pp. 508–514.
24. G. V. Trunk and P. K. Hughes II, "Automatic detectors for frequency-agile radar," in *IEE Int. Radar Conf*., London, 1982, pp. 464–468.
25. G. V. Trunk, B. H. Cantrell, and F. D. Queen, "Modified generalized sign test processor for 2-D radar," *IEEE Trans*., vol. AES-10, pp. 574-582, September 1974.
26. J. T. Rickard and G. M. Dillard, "Adaptive detection algorithms for multiple-target situations," *IEEE Trans*., vol. AES-13, pp. 338–343, July 1977.
27. H. M. Finn, "A CFAR design for a window spanning two clutter fields," *IEEE Trans*., vol. AES-22, pp. 155–168, March 1986.
28. B. A. Green, "Radar detection probability with logarithmic detectors," *IRE Trans*., vol. IT-4, March 1958.
29. V. G. Hansen and J. R. Ward, "Detection performance of the cell average log/CFAR receiver," *IEEE Trans*., vol. AES-8, pp. 648–652, September 1972.
30. G. M. Dillard and C. E. Antoniak, "A practical distribution-free detection procedure for multiple-range-bin radars," *IEEE Trans*., vol. AES-6, pp. 629–635, September 1970.
31. V. G. Hansen and B. A. Olsen, "Nonparametric radar extraction using a generalized sign test," *IEEE Trans*., vol. AES-7, September 1981.
32. W. G. Bath, L. A. Biddison, S. F. Haase, and E. C. Wetzlar, "False alarm control in automated radar surveillance systems," in *IEE Int. Radar Conf*., London, 1982, pp. 71–75.
33. C. E. Muehe, L. Cartledge, W. H. Drury, E. M. Hofstetter, M. Labitt, P. B. McCorison, and V. J. Sferrino, "New techniques applied to air-traffic control radars," *Proc. IEEE*, vol. 62, pp. 716–723, June 1974.
34. G. V. Trunk, "Range resolution of targets using automatic detectors," *IEEE Trans*., vol. AES-14, pp. 750–755, September 1978.
35. G. V. Trunk, "Range resolution of targets," *IEEE Trans*., vol. AES-20, pp. 789–797, November 1984.
36. G.V. Trunk and S. M. Brockett, "Range and velocity ambiguity resolution," in *IEEE National Radar Conf*., Boston, 1993, pp. 146–149.
37. G.V. Trunk and M. Kim, "Ambiguity resolution of multiple targets using pulse-doppler waveforms," *IEEE Trans*., vol. AES-30, pp. 1130-1137, October 1994.
38. H. Leung, Z. Hu, and M. Blanchette, "Evaluation of multiple radar target trackers in stressful environments," *IEEE Trans*. *Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 663–674, 1999.
39. B. H. Cantrell, G. V. Trunk, and J. D. Wilson, "Tracking system for two asynchronously scanning radars," Naval Res. Lab. Rept. 7841, 1974.
40. W. D. Stuckey, "Activity control principles for automatic tracking algorithms," in *IEEE Radar 92 Conference*, 1992, pp. 86–89.
41. T. R. Benedict and G. W. Bordner, "Synthesis of an optimal set of radar track-while-scan filtering equations," *IRE Trans*., vol. AC-7, pp. 27–32, 1962.
42. R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng*. (*ASME Trans*., ser. D), vol. 82, pp. 35–45, 1960.
43. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *J. Basic Eng*. (*ASME Trans*., ser. D), vol. 83, pp. 95–107, 1961.

44. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Boston: Artech, 1999.
45. R. A. Singer, "Estimating optimal tracking filter performance for manned maneuvering targets," *IEEE Trans*., vol. AES-6, pp. 472–484, 1970.
46. B. Friedland, "Optimum steady state position and velocity estimation using noisy sampled position data," *IEEE Trans*. vol. AES, p. 906, 1973.
47. P. Kalata, "The tracking index: A generalized parameter for $\alpha - \beta$ and α, β, γ target trackers," IEEE Trans. Aerospace and Electronic Systems, AES-20, pp. 174–182, 1984.
48. W. D. Blair and Y. Bar-Shalom, "Tracking maneuvering targets with multiple sensors: Does more data always mean better estimates?" *IEEE Trans. Aerospace and Electronic Systems,* vol. 32, pp. 450-456, 1996.
49. F. R.Castella, "Analytical results for the x,y Kalman tracking filter," *IEEE Trans. Aerospace and Electronic Systems*, November 1974, vol. 10, pp.891-894,
50. R. F. Fitzgerald, "Simple tracking filters: Steady-state filtering and smoothing performance," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-16, pp. 860–864, 1980.
51. G. J. Portmann, J. Moore, and W. G. Bath, "Separated covariance filtering," in *Rec. IEEE 1990 International Radar Conference*, 1990, pp. 456–460.
52. P. Mookerjee and F. Reifler, "Reduced state estimator for systems with parametric inputs," *IEEE Trans. Aerospace and Electronic Systems*, vol. 40, no. 2, pp. 446–461, 2004.
53. A. S. Gelb, *Applied Optimal Estimation*, Cambridge, MA: MIT Press, 1974.
54. F. R. Castella, "Multisensor, multisite tracking filter," *IEE Proc. Radar, Sonar Navigation*, vol. 141, issue 2, pp. 75–82, 1994.
55. E. A. Wan, R. van der Merwe, and A. T. Nelson, "Dual estimation and the unscented transformation," in *Advances in Neural Information Processing Systems 12*, Cambridge: MIT Press, 2000, pp. 666–672.
56. G. A. Watson and W. D. Blair, "IMM algorithm for tracking targets that maneuver through coordinated turns," *SPIE, Signal and Data Processing of Small Targets*, vol. 1698, pp. 236–247, 1992.
57. R. Cooperman, "Tactical ballistic missile tracking using the interacting multiple model algorithm," in *Proc. Fifth International Conference on Information Fusion*, vol. 2, 2002, pp. 824–831.
58. C. L. Morefield, "Application of 0–1 integer programming to multi-target tracking problems," *IEEE Trans*., vol. AC-22, pp. 302–312, 1977.
59. R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, vol. 38, no. 4, pp. 325–340, 1987.
60. D. Bertsekas, "The auction algorithm for assignment and other network flow problems: A tutorial," *Interfaces*, vol. 20, pp. 133–149, 1990.
61. I. Kadar, E. Eadan, and R. Gassner, "Comparison of robustized assignment algorithms," *SPIE*, vol. 3068, pp. 240–249, 1997.
62. Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, pp. 451–460, 1975.
63. S. B. Colegrove and J. K. Ayliffe, "An extension of probabilistic data association to include track initiation and termination," in *20th IREE Int. Conv. Dig.*, Melbourne, Austrailia, 1985, pp. 853–856.
64. S. B. Colegrove, A. W. Davis, and J. K. Ayliffe, "Track initiation and nearest neighbors incorporated into probabilistic data association," *J. Elec. Electron. Eng. (Australia), IE Aust*. and *IREE Aust*., vol. 6, pp. 191–198, 1986.
65. Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*, Orlando, FL: Academic Press, 1988.
66. R. W. Sittler, "An optimal association problem in surveillance theory," *IEEE Trans*., vol. MIL-8, pp. 125–139, 1964.
67. J. J. Stein and S. S. Blackman, "Generalized correlation of multi-target track data," *IEEE Trans*., vol. AES-11, pp. 1207–1217, 1975.

68. G. V. Trunk and J. D. Wilson, "Track initiation of occasionally unresolved radar targets," *IEEE Trans*., vol. AES-17, pp. 122–130, 1981.
69. W. Koch, "On Bayesian MHT for well separated targets in densely cluttered environment," in *Proc. IEEE International Radar Conference*, 1995, pp. 323–328.
70. D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter," *SPIE, Signal and Data Processing of Small Targets*, vol. 1305, pp. 434–445, 1990.
71. R. J. Prengaman, R. E. Thurber, and W. G. Bath, "A retrospective detection algorithm for extraction of weak targets in clutter and interference environments," in *IEEE Int. Radar Conf*., London, 1982, pp. 341–345.
72. N. Levine, "A new technique for increasing the flexibility of recursive least squares smoothing," *Bell System Technical Journal*, pp. 819–840, 1961.
73. W. G. Bath, M. E. Baldwin, and W. D. Stuckey, "Cascaded spatial correlation processes for dense contact environments," in *Proc. RADAR 1987*, 1987, pp. 125–129.
74. R. J. Prengaman, R. E. Thurber, and W. G. Bath, "A retrospective detection algorithm for extraction of weak targets in clutter and interference environments," in *IEEE Int. Radar Conf*., London, 1982, pp. 341–345.
75. E. R. Billam, "Parameter optimisation in phased array radar," in *Radar 92*, Brighton, UK, 12–13 October 1992, pp. 34–37.
76. G. V. Trunk, J. D. Wilson, and P. K. Hughes, II, "Phased array parameter optimization for low-altitude targets," in IEEE 1995 International Radar Conference, May 1995 pp. 196–200.
77. W. Bath, "Tradeoffs in radar networking," in *Proc. IEE RADAR 2002*, 2002, pp. 26–30.
78. J. R. Moore and W. D. Blair, "Practical aspects of multisensor tracking," in *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. III, Boston: Artech House, 2000.
79. J. O. Coleman, "Discriminants for assigning passive bearing observations to radar targets," in *IEEE Int. Radar Conf*., Washington, DC, 1980, pp. 361–365.
80. G. V. Trunk and J. D. Wilson, "Association of DF bearing measurements with radar tracks," *IEEE Trans*., vol. AES-23, 1987, pp. 438–447.
81. G. V. Trunk and J. D. Wilson, "Correlation of DF bearing measurements with radar tracks," in *IEEE Int. Radar Conf*., London, 1987, pp. 333–337.

Chapter 8
Pulse Compression Radar

Michael R. Ducoff
Byron W. Tietjen
Lockheed Martin MS2

8.1 INTRODUCTION*

A pulse compression radar transmits a long pulse with pulsewidth T and peak power P_t, which is coded using frequency or phase modulation to achieve a bandwidth B that is large compared to that of an uncoded pulse with the same duration.[1] The transmit pulsewidth is chosen to achieve the single-pulse transmit energy, given by $E_{t1} = P_t T$, that is required for target detection or tracking. The received echo is processed using a pulse compression filter to yield a narrow compressed pulse response with a mainlobe width of approximately 1/B that does not depend on the duration of the transmitted pulse.

Figure 8.1 shows a block diagram of a basic pulse compression radar. The coded pulse is generated at a low power level in the waveform generator and amplified to the required peak transmit power using a power amplifier transmitter. The received signal is mixed to an intermediate frequency (IF) and amplified by the IF amplifier. The signal is then processed using a pulse compression filter that consists of a matched filter to achieve maximum signal-to-noise ratio (SNR). As discussed below, the matched filter is followed by a weighting filter if required for reduction of time sidelobes. The output of the pulse compression filter is applied to an envelope detector, amplified by the video amplifier, and displayed to an operator.

The ratio of the transmit pulsewidth to the compressed pulse mainlobe width is defined as the pulse compression ratio. The pulse compression ratio is approximately $T/(1/B)$ or TB, where TB is defined as the time-bandwidth product of the waveform. Typically, the pulse compression ratio and time-bandwidth product are large compared to unity.

The use of pulse compression provides several performance advantages. The increased detection range capability of a long-pulse radar system is achieved with pulse compression while retaining the range resolution capability of a radar that uses a narrow uncoded pulse. The required transmitted energy can be established by

* The authors would like to acknowledge the use of material previously prepared by Edward C. Farnett and George H. Stevens for the "Pulse Compression Radar" chapter in the second edition of the *Radar Handbook* (1990), edited by Merrill I. Skolnik.

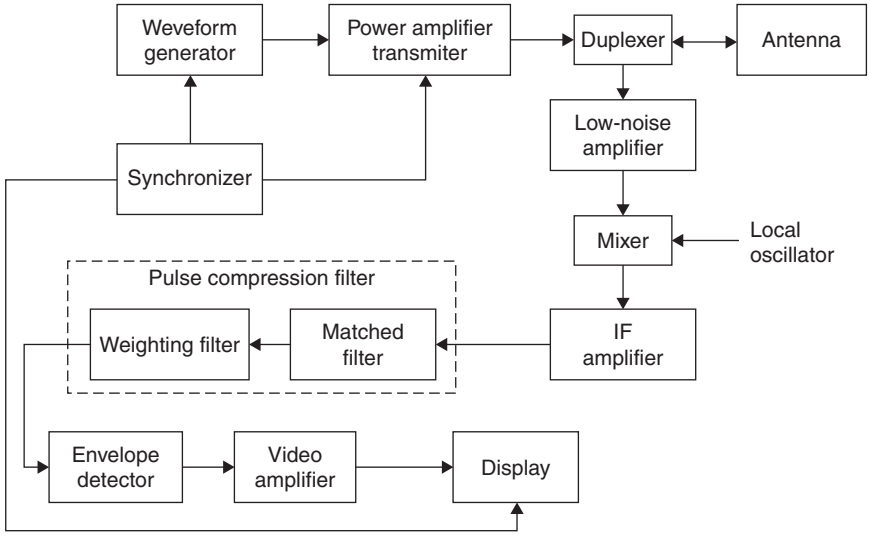


FIGURE 8.1 Block diagram of a basic pulse compression radar

increasing the waveform pulsewidth without exceeding constraints on transmitter peak power. The average power of the radar may be increased without increasing the pulse repetition frequency (PRF) and, hence, decreasing the radar's unambiguous range. In addition, the radar is less vulnerable to interfering signals that differ from the coded transmitted signal.

The mainlobe of the compressed pulse at the output of the matched filter has time, or range, sidelobes that occur within time intervals of duration T, before and after the peak of the peak of the compressed pulse. The time sidelobes can conceal targets, which would otherwise be resolved using a narrow uncoded pulse. In some cases, such as phase-coded waveforms or nonlinear frequency modulation waveforms, matched filter processing alone achieves acceptable time sidelobe levels. However, for the case of a linear frequency modulation waveform, the matched filter is generally followed by a weighting filter to provide a reduction in time sidelobe levels. In this case, the weighting filter results in a signal-to-noise ratio loss compared to that of matched filter processing alone.

8.2 PULSE COMPRESSION WAVEFORM TYPES

The following sections describe the characteristics of the linear and nonlinear frequency modulation waveforms, phase-coded waveforms, and time-frequency coded waveforms. The application of surface acoustic wave (SAW) devices for linear frequency modulation (LFM) waveform pulse compression is discussed. Waveform signal analysis techniques, matched filter properties, and the waveform autocorrelation and ambiguity function definitions used are summarized in the Appendix at the end of this chapter.

Linear Frequency Modulation.[1,2] The linear frequency modulation, or chirp, waveform has a rectangular amplitude modulation with pulsewidth T and a linear frequency modulation with a swept bandwidth B applied over the pulse. The time-bandwidth product of the LFM waveform is equal to TB, where TB is the product of pulsewidth and swept bandwidth. The 3-dB width of the compressed pulse at the output of the matched filter is $\tau_3 = 0.886/B$, for large values of time-bandwidth product. The peak time sidelobe level of the compressed pulse is –13.2 dB.

As discussed in Section 8.1, a frequency-domain weighting filter is generally required following the matched filter to provide reduced time sidelobe levels, at the cost of reduced SNR and an increase in the width of the compressed pulse. As an example, the use of 40-dB Taylor weighting reduces the peak time sidelobe level from –13.2 dB to –40 dB and introduces a 1.15 dB loss in SNR. The 3-dB width of the compressed pulse with weighting increases from $\tau_3 = 0.886/B$ to $\tau_3 = 1.25/B$.

The LFM waveform has a knife-edge ambiguity function with contours that are approximately elliptical with a major axis defined by the line $v = \alpha\tau$, where $\alpha = \pm B/T$ is the LFM slope. This property introduces range-doppler coupling at the matched filter output causing the matched filter output peak to occur earlier in time for a target with a positive doppler frequency compared to a stationary target at the same range, assuming a positive linear frequency modulation slope and later in time for a negative slope.

The compressed pulse shape and SNR are tolerant to doppler shift for the LFM waveform. As a result, it is not necessary to implement multiple matched filters to cover the range of expected target doppler shifts.

LFM Waveform Definition. The LFM waveform is a single-pulse bandpass signal defined as

$$x(t) = A \text{ rect } (t/T) \cos [2\pi f_0 t + \pi\alpha t^2] \tag{8.1}$$

where T is the pulsewidth, f_0 is the carrier frequency, α is the LFM slope, and the rect function is defined as

$$\text{rect}(x) = \begin{cases} 1, |x| < 1/2 \\ 0, |x| > 1/2 \end{cases} \tag{8.2}$$

The LFM slope is given by $\alpha = \pm B/T$, where the plus sign applies for a positive LFM slope (termed an *up-chirp*) and the minus sign for a negative LFM slope (a *down-chirp*). The amplitude modulation is $a(t) = A \text{ rect}(t/T)$ and the phase modulation is a quadratic function of time:

$$\phi(t) = \pi\alpha t^2 \tag{8.3}$$

The frequency modulation, defined as the instantaneous frequency deviation from the carrier frequency f_0, is expressed in terms of the phase modulation by

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi}{dt} \tag{8.4}$$

The frequency modulation for an LFM waveform is linear with slope equal to α

$$f_i(t) = \alpha t = \pm(B/T)t, \quad |t| \le T/2 \tag{8.5}$$

where the plus sign applies for a positive LFM slope and the minus sign for a negative slope. The complex envelope of the LFM waveform is expressed in terms of the amplitude and phase modulation functions as

$$u(t) = A \text{ rect}(t/T)e^{j\pi\alpha t^2}$$

Figure 8.2 shows an example of an LFM bandpass signal with a pulsewidth $T = 10$ μs, swept bandwidth $B = 1$ MHz and time-bandwidth product equal to TB = 10. The LFM slope is $B/T = 0.1$ MHz/μs. The instantaneous frequency of the LFM waveform varies between 1.5 and 2.5 MHz over the pulse duration, as indicated by the reduction in the spacing of successive positive-going zero crossings of the signal.†

LFM Waveform Spectrum.[1,2,3] The spectrum of the LFM waveform has a significant amplitude variation versus frequency for small time-bandwidth products. For large values of time-bandwidth product, the magnitude of the spectrum approaches rect(f/B)

$$u(t) = \frac{1}{\sqrt{T}} \text{rect}(t/T)e^{j\pi\alpha t^2}$$
$$|U(f)| \approx \text{rect}(f/B) \quad \text{for TB} >> 1 \tag{8.6}$$

The LFM spectrum is expressed in terms of the complex Fresnel integral, and the amplitude variation present for low values of TB is termed the *Fresnel ripple*.

LFM Waveform Ambiguity Function. The waveform autocorrelation function and ambiguity function for an LFM waveform are given by

$$\chi_u(\tau, f_d) = [1 - |\tau/T|] \text{ sinc}[(f_d - \alpha\tau)T(1 - |\tau/T|)] \text{ rect } (\tau/2T)e^{-j\pi f_d \tau} \tag{8.7}$$

$$\Psi_u(\tau, f_d) = [1 - |\tau/T|]^2 \text{ sinc}^2[(f_d - \alpha\tau)T(1 - |\tau/T|)] \text{ rect}(\tau/2T) \tag{8.8}$$

where the sinc function is defined as

$$\text{sinc}(x) = \sin(\pi x)/(\pi x)$$

The matched filter time response for a target with doppler shift f_d is obtained by the substitution $\tau = -t$ in the autocorrelation function:

$$y(t) = \chi_u(-t, f_d) = [1 - |t/T|] \text{ sinc}[(f_d + \alpha t)T(1 - |t/T|)] \text{ rect}(t/2T)e^{j\pi f_d t} \tag{8.9}$$

LFM Range-doppler Coupling. The LFM waveform exhibits range-doppler coupling which causes the peak of the compressed pulse to shift in time by an amount proportional to the doppler frequency. The peak occurs earlier in time at $t = -f_d T/B$ for a positive LFM slope, compared to peak response for a stationary target. The peak of the ambiguity function is shifted to $\tau = f_d T/B$ for a positive LFM slope.

Time Delay and Range Resolution Widths. The time-delay resolution width is equal to the width of the ambiguity function at a specified level relative to the peak value.

† Low values of carrier frequency and time-bandwidth product have been used to illustrate the variation of instantaneous frequency over the pulse in Figure 8.2.

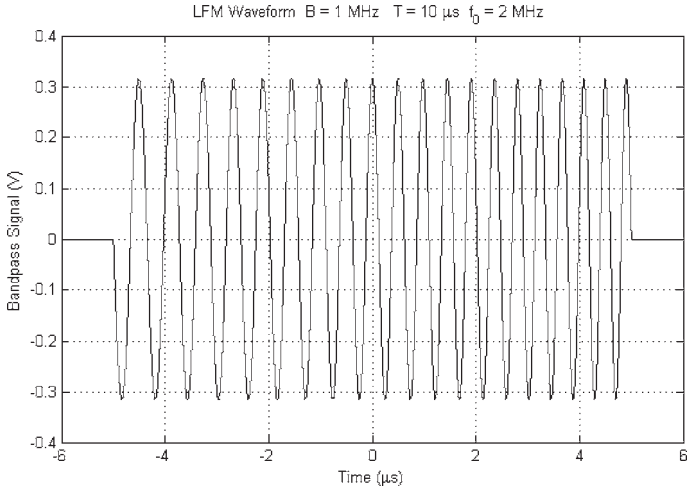


FIGURE 8.2 LFM bandpass signal example (shown for T = 10 μs, B = 1 MHz, f_0 = 2 MHz)

For the case of a large time-bandwidth, the magnitude of the autocorrelation function measured along the relative time delay axis is given by

$$|\chi_u(\tau,0)| \approx |\text{sinc}(B\tau)|, \quad |\tau| \ll T$$

The x-dB time delay resolution is measured between the values of τ for which

$$20\log|\text{sinc}(B\tau)| = -x(\text{dB})$$

The range resolution is equal to $c/2$ times the corresponding time delay resolution where c is the speed of light. Table 8.1 contains a summary of the resolution widths for the LFM waveform.

LFM Waveform Examples. Figure 8.3 shows the magnitude of the autocorrelation function as a function of relative time delay τ for doppler shifts‡ of –0.5 MHz, 0 and 0.5 MHz, pulsewidth T = 10 μs, swept bandwidth B = 1 MHz, and LFM slope $\alpha = B/T$ = 0.1 MHz/μs. A doppler shift of $f_d = B/2$ = 0.5 MHz causes the peak of the correlation function to move to $\tau = f_d T/B$ = 5 μs. Figure 8.4 shows the result when the pulsewidth is increased to 100 μs to yield a waveform with an LFM slope equal

TABLE 8.1 LFM Waveform Time Delay and Range Resolution Widths

Mainlobe Width	Time Delay Resolution (s)	Range Resolution (m)
3.01 dB	$\tau_3 = 0.886/B$	$\Delta R_3 = 0.886c/B$
3.9 dB	$\tau_{3.9} = 1/B$	$\Delta R_{3.9} = c/2B$
6.02 dB	$\tau_6 = 1.206/B$	$\Delta R_6 = 1.206c/2B$
10.0 dB	$\tau_{10} = 1.476/B$	$\Delta R_{10} = 1.476c/2B$

‡ These values of doppler shift are large for microwave radars and were selected to show the effect of range-doppler coupling.

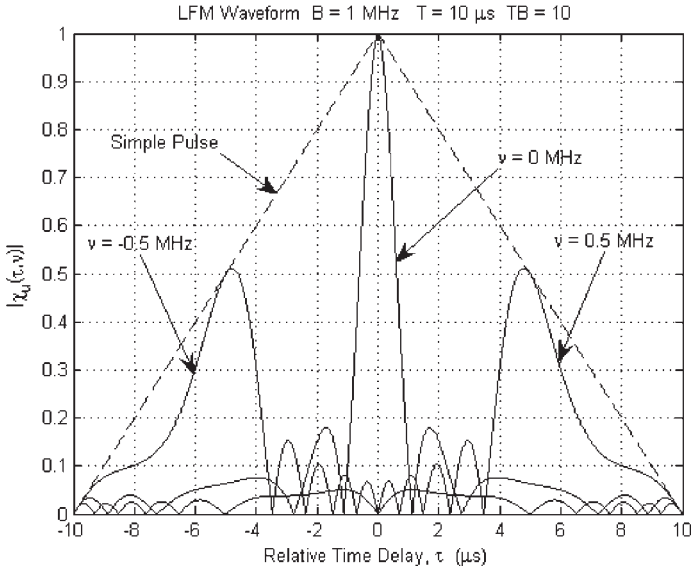


FIGURE 8.3 LFM waveform autocorrelation function (T = 10 μs, B = 1 MHz, TB = 10)

to 0.01 MHz/μs. In this case, a doppler shift of 0.5 MHz shifts the peak of autocorrelation function to $\tau = 50$ μs, an increase of a factor of ten compared to the result for a 10-μs pulsewidth.

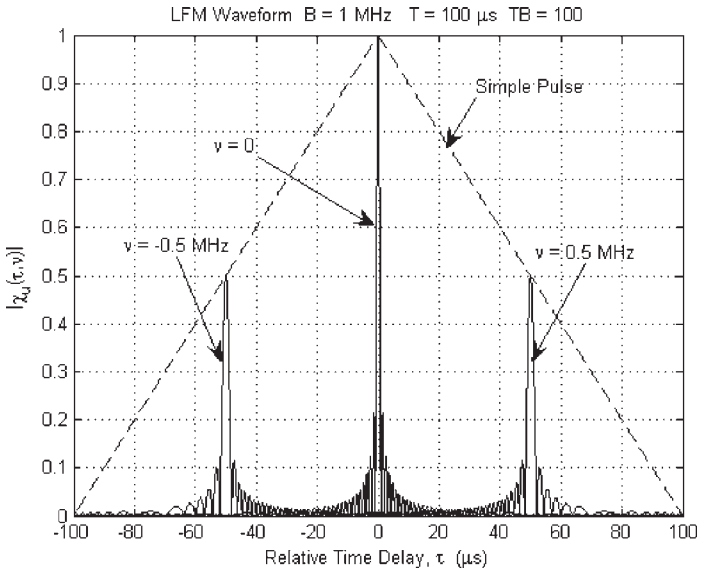


FIGURE 8.4 LFM waveform autocorrelation function (T = 100 μs, B = 1 MHz, TB = 100)

Frequency Domain Weighting for LFM Time Sidelobe Reduction.[1,2,4] A frequency domain weighting filter is used following the matched filter for time sidelobe reduction. Taylor weighting provides a realizable approximation to the ideal Dolph-Chebyshev weighting, which achieves the minimum mainlobe width for a given value of peak time sidelobe level. The frequency response of the equivalent low pass filter for the Taylor weighing filter is

$$W(f) = 1 + 2\sum_{m=1}^{\bar{n}-1} F_m \cos\left(2\pi \frac{mf}{B}\right) \tag{8.10}$$

where F_m is the Taylor coefficient and $\bar{n}$ is the number of terms in the weighting function. The compressed pulse response at the output of the weighting filter is given by

$$y_o(t) = \text{sinc}(Bt) + \sum_{m=1}^{\bar{n}-1} F_m[\text{sinc}(Bt + m) + \text{sinc}(Bt - m)] \tag{8.11}$$

As discussed below, the compressed pulse response (Eq. 8.11) is based on the assumption that the time-bandwidth product of the LFM waveform is much greater than unity ($TB \gg 1$). The filter matching loss for Taylor weighting is given by Klauder et al.[1] as

$$L_m = 1 + 2\sum_{m=1}^{\bar{n}-1} F_m^{\,2} \tag{8.12}$$

Figure 8.5 shows a comparison of the compressed pulse response for three frequency domain weighting types: Curve A is for uniform weighting where $W(f) = 1$

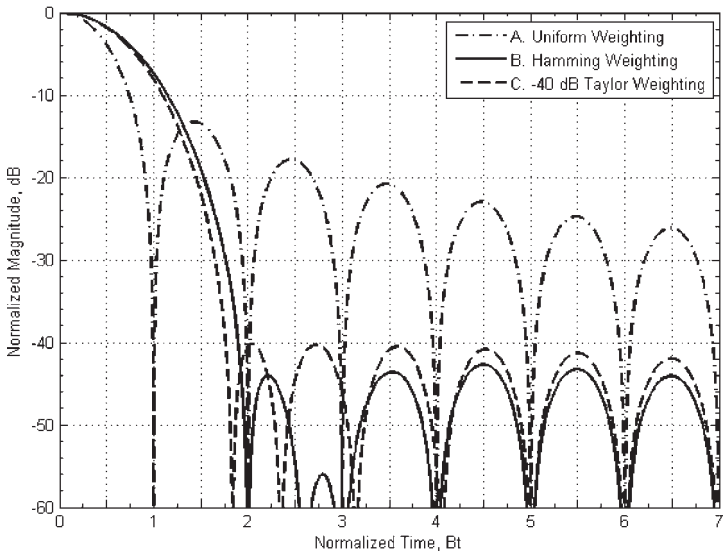


FIGURE 8.5 Comparison of compressed-pulse shapes for three frequency-domain weighting functions

(matched filter processing); Curve C is for Taylor weighting with –40 dB peak time sidelobe level ($\overline{n} = 6$); and Curve B is for Hamming weighting where

$$W(f) = 1 + 2F_1 \cos\left(2\pi \frac{mf}{B}\right)$$

$$F_1 = 0.4259 \tag{8.13}$$

The Taylor coefficients for –40-dB Taylor weighting ($\overline{n} = 6$) are listed here:[5]

$F_1 = 0.389116$
$F_2 = -0.00945245$
$F_3 = 0.00488172$
$F_4 = -0.00161019$
$F_5 = 0.000347037$

Table 8.2 shows the peak time sidelobe level, 3-dB and 6-dB compressed-pulse widths, and filter matching loss for the three weighting function types. The application of –40-dB Taylor weighting reduces the peak time sidelobe level from –13.2 dB to –40 dB and increases the filter matching loss from 0 dB to 1.15 dB. The 3-dB compressed-pulse mainlobe width increases from $0.886/B$ to $1.25/B$ when –40-dB Taylor weighting is used. The 3-dB and 6-dB mainlobe widths and filter matching loss for Hamming weighting are approximately the same as for –40-dB Taylor weighting.

These results assume that the time-bandwidth product of the LFM waveform is much greater than unity so that the time sidelobe performance is not limited by the Fresnel amplitude ripple in the spectrum of the LFM waveform. Cook and Paolillo[3] and Cook and Bernfeld[2] have analyzed the effect of the Fresnel amplitude ripple and pulse rise-time and fall-time on time sidelobe levels. A phase predistortion technique is described by Cook and Paolillo,[3] which reduces the Fresnel amplitude ripple to allow low time sidelobes to be achieved for LFM waveforms with relatively small time-bandwidth products.

Radar equipment distortion sources also establish limitations on achievable time sidelobe levels and are discussed by Klauder et al.[1] and Cook and Bernfeld.[2] The method of paired-echo analysis is used to evaluate the effect of amplitude and phase distortion on the time sidelobe levels. Frequency domain amplitude and phase distortion is typically caused by filters and transmission line reflections. Time domain amplitude and phase distortion, termed *modulation distortion* by Cook and Bernfeld, can result from power supply ripple in high-power transmitter amplifiers.[2]

TABLE 8.2 Comparison of LFM Weighting Filters

Weighting Function	Peak Time Sidelobe Level (dB)	3-dB Mainlobe Width, τ_3	6-dB Mainlobe Width, τ_3	Filter Matching Loss (dB)
Uniform	–13.2	$0.886/B$	$1.21/B$	0
Taylor (–40 dB, $\overline{n} = 6$)	–40	$1.25/B$	$1.73/B$	1.15
Hamming	–43	$1.30/B$	$1.81/B$	1.34

Taylor Versus Cosine-Squared-Plus-Pedestal Weighting. Figure 8.6*a* plots the taper coefficient F_1 and pedestal height H versus the peak time sidelobe level for cosine-squared-plus-pedestal weighting. For a given peak time sidelobe level, Taylor weighting offers theoretical advantages in range resolution and SNR performance, as illustrated in Figure 8.6*b* and Figure 8.6*c*.

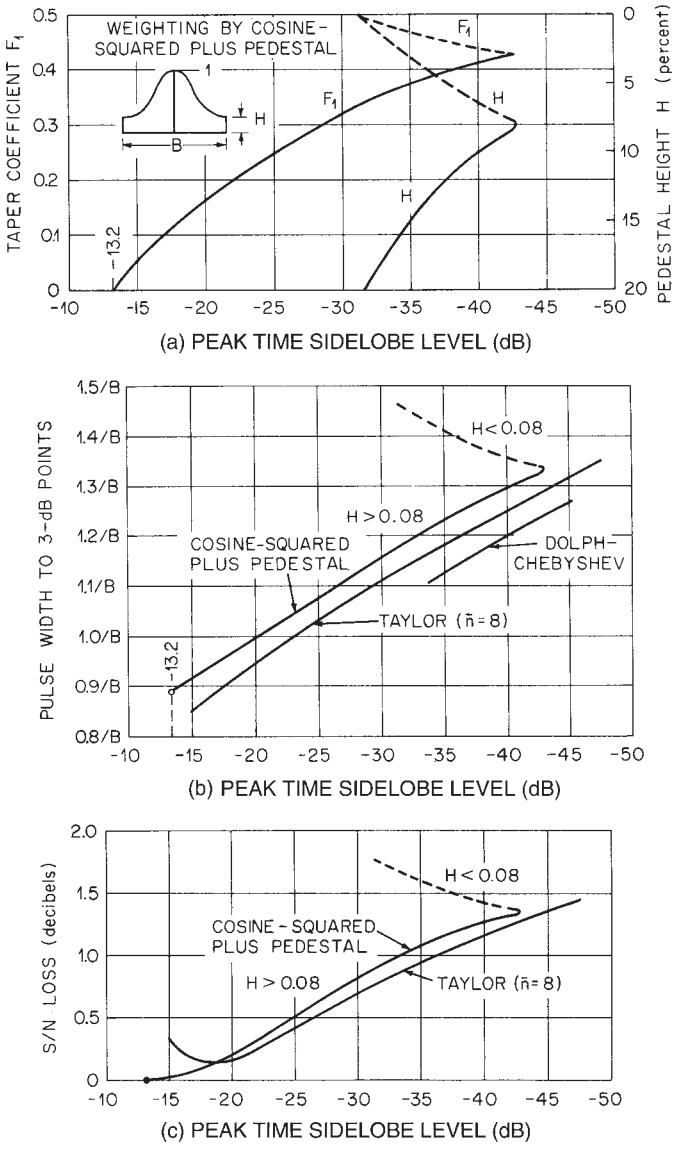


FIGURE 8.6 (*a*) Taper coefficient and pedestal height versus peak time sidelobe level; (*b*) Compressed-pulse width versus peak time sidelobe level; (*c*) *SNR* loss versus peak time sidelobe level

SAW Devices for LFM Pulse Compression. A Surface Acoustic Wave (SAW) device consists of an input transducer and an output transducer mounted on a piezoelectric substrate. These transducers are usually implemented as interdigital devices that consist of a metal film deposited on the surface of the acoustic medium. This metal film is made of fingers (see Figure 8.7) that dictate the frequency characteristic of the unit. The input transducer converts an electrical signal into a sound wave with over 95% of the energy traveling along the surface of the medium. The output transducer taps a portion of this surface sound wave and converts it back into an electric signal.

The SAW device[6-8] has unique features that dictate its usefulness for a given radar application. It represents one of the few analog processing devices used in modern radar. The advantages of the SAW device are its compact size, the wide bandwidths that can be attained, the ability to tailor the transducers to a particular waveform, the all-range coverage of the device, and the low cost of reproducing a given design. The major shortcomings of the SAW approach are that the waveform length is restricted. Since sound travels about 3 to 15 mm/μs on the surface of a SAW device, a 250 mm quartz device (about the largest available), has a usable delay of about 70 μs for a single pass.[9] Also, because each SAW device is waveform specific, each waveform requires a different design.

SAW pulse compression devices depend on the interdigital transducer finger locations or the surface-etched grating to determine its bandpass characteristics. Figure 8.7 shows three types of filter determination approaches. A wideband input transducer and a frequency-selective (dispersive) output transducer are used in Figure 8.7*a*. When an impulse is applied to the input, the output signal is initially a low frequency that increases (based on the output transducer finger spacings) at later portions of the pulse. This results

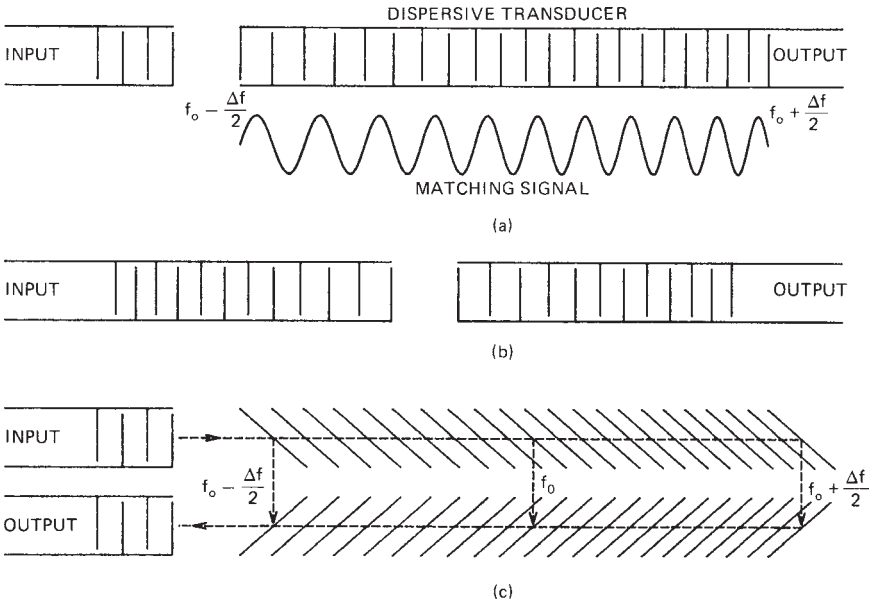


FIGURE 8.7 SAW transducer types: (*a*) dispersive output, (*b*) both input and output dispersive, and (*c*) dispersive reflections

in an up-chirp waveform that would be a matched filter for a down-chirp transmitted waveform. In Figure 8.7*b* both the input transducer and the output transducer are dispersive, which would result in the same impulse response as that shown in Figure 8.7*a*. For a given crystal length and material, the waveform duration for the approaches in Figure 8.7*a* and Figure 8.7*b* would be the same and is limited to the time that it takes an acoustic wave to traverse the crystal length. Figure 8.7*c* shows a reflection-array-compression (RAC) approach[10] that essentially doubles the achievable pulse length for the same crystal length. In an RAC, the input and output transducers have a broad bandwidth. A frequency-sensitive grating is etched on the crystal surface to reflect a portion of the surface-wave signal to the output transducer. This grating coupling does not have a significant impact on the surface-wave energy. Except for a 2:1 increase in the waveform duration, the impulse response of the RAC is the same as for the approaches shown in Figure 8.7*a* and *b*. Thus, these three approaches yield a similar impulse response.

Figure 8.8 shows a sketch of a SAW pulse compression device with dispersive input and output transducers. As the energy in a SAW device is concentrated in its surface wave, the SAW approach is much more efficient than bulk-wave devices, where the wave travels through the crystal. The propagation velocity of the surface wave is in the range of 1500 to 4000 m/s, depending on the crystal material, and allows a large delay in a compact device. Acoustic absorber material is required at the crystal edges to reduce the reflections and, hence, the spurious responses. The upper frequency limit depends on the accuracy that can be achieved in the fabrication of the interdigital transducer. The SAW device must provide a response that is centered on a carrier, as the lowest frequency of operation is about 20 MHz and is limited by the crystal. A matched-filter SAW pulse compression device can use variable finger lengths to achieve frequency weighting, and this internal weighting can correct for the Fresnel amplitude ripples[11] in the FM spectrum. With this correction, –43 dB time sidelobe levels can be achieved for a linear-FM waveform with TB as low as 15. The level of sidelobe suppression depends upon the time bandwidth product, the weighting function applied, and fabrication errors in the SAW device. Time sidelobe levels of –35 dB have been achieved for TB between 5 and 15. TB products of up to 2000 have been achieved[12] with time sidelobes better than –40 dB.[13] Dynamic range is limited by nonlinearities in the crystal material, but dynamic ranges over 90 dB have been achieved. The most common SAW materials are quartz, lithium niobate, and lithium tantalite.

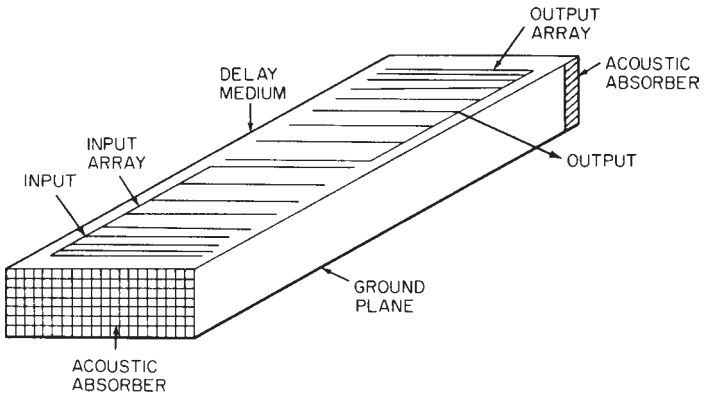


FIGURE 8.8 Surface-wave delay line

Nonlinear Frequency Modulation Waveforms. The nonlinear-FM waveform has several distinct advantages over LFM.[14,16] It requires no frequency domain weighting for time sidelobe reduction because the FM modulation of the waveform is designed to provide the desired spectrum shape that yields the required time sidelobe level. This shaping is accomplished by increasing the rate of change of frequency modulation near the ends of the pulse and decreasing it near the center. This serves to taper the waveform spectrum so that the matched filter response has reduced time sidelobes.[16] Thus, the loss in signal-to-noise ratio associated with frequency domain weighting (as for the LFM waveform) is eliminated.

If a symmetrical FM modulation is used (Figure 8.9*a*) with time-domain amplitude weighting to reduce the frequency sidelobes, the nonlinear-FM waveform will have a thumbtack-like ambiguity function (Figure 8.10). A symmetrical waveform typically has a frequency that increases (or decreases) with time during the first half of the pulse and decreases (or increases) during the last half of the pulse. A nonsymmetrical waveform is obtained by using one-half of a symmetrical waveform (Figure 8.9*b*). However, the nonsymmetrical waveform retains some of the range-doppler coupling of the linear-FM waveform.

One of the primary disadvantages of the nonlinear-FM waveform is that it is less doppler tolerant than the LFM. In the presence of doppler shift, the time sidelobes of the pulse-compressed NLFM tend to increase compared to those of the LFM. Figure 8.14, shown later in this section, and Table 8.3 illustrate this behavior for a typical NLFM pulse.

This characteristic of the NLFM waveform sometimes necessitates processing using multiple matched filters offset in doppler shift to achieve the required time sidelobe level. Because of the doppler sensitivity of the ambiguity function, the nonlinear frequency modulation waveform is useful in a tracking system where range and doppler frequency are approximately known, and the target doppler shift can be compensated in the matched filter. The nonsymmetrical NLFM waveform is used in the MMR system, for example, which detects and tracks ordnance such as mortars, artillery, and rockets.

To achieve a –40-dB Taylor compressed pulse response, for example, the frequency-versus-time (frequency modulation) function of a nonsymmetrical NLFM waveform of bandwidth B is[14]

$$f(t) = B\left(\frac{t}{T} + \sum_{n=1}^{7} K_n \sin\frac{2\pi nt}{T}\right) \tag{8.14}$$

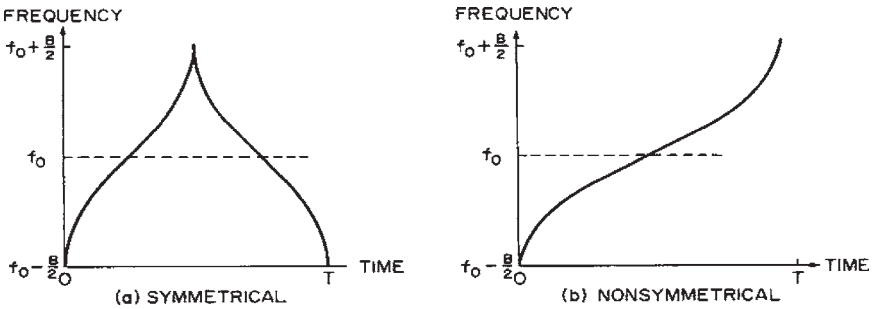


FIGURE 8.9 Symmetrical and nonsymmetrical nonlinear-FM waveforms

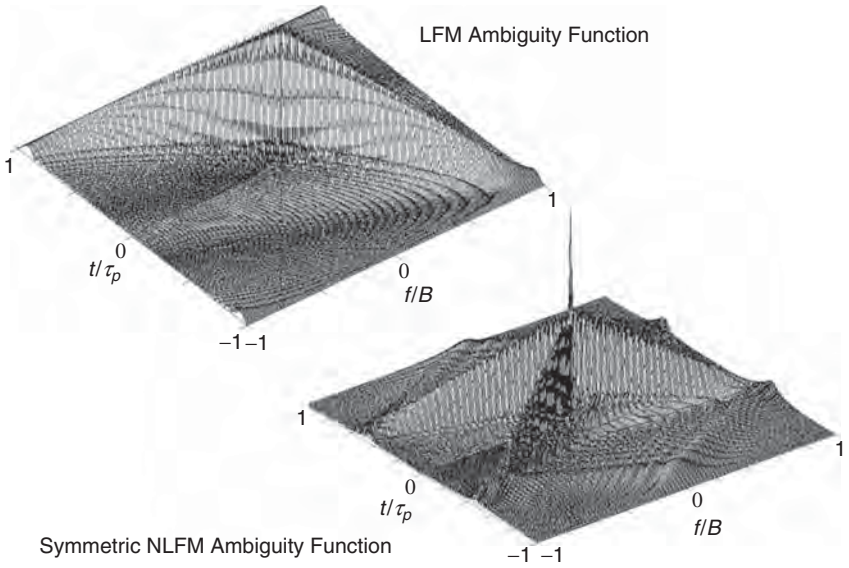


FIGURE 8.10 Ambiguity function of an LFM waveform compared to a symmetrical NLFM waveform

where the coefficients are

$K_1 = -0.1145$
$K_2 = 0.0396$
$K_3 = -0.0202$
$K_4 = 0.0118$
$K_5 = 0.0082$
$K_6 = 0.0055$
$K_7 = -0.0040$

Other NLFM waveforms that have been utilized in radar include the nonsymmetrical sine-based and tangent-based waveforms.[§] For the sine-based waveform, the relationship between time and frequency modulation is given as

$$\frac{t}{T} = \frac{f}{B} + \frac{k}{2\pi}\sin(2\pi f/B) \quad \text{for } -B/2 \le f \le B/2 \tag{8.15}$$

where T is the pulsewidth, B is the swept bandwidth, and k is a time sidelobe level control factor.

Typical k values are 0.64 and 0.70, which yield time sidelobe levels of –30 dB and –33 dB, respectively. Figure 8.11 is a plot of peak time sidelobe level as a function of the time sidelobe control factor k, for various TB products, for this NLFM waveform.

[§] Courtesy of Edwin M. Waterschoot, Lockheed Martin Maritime and Sensor Systems, Syracuse, NY.

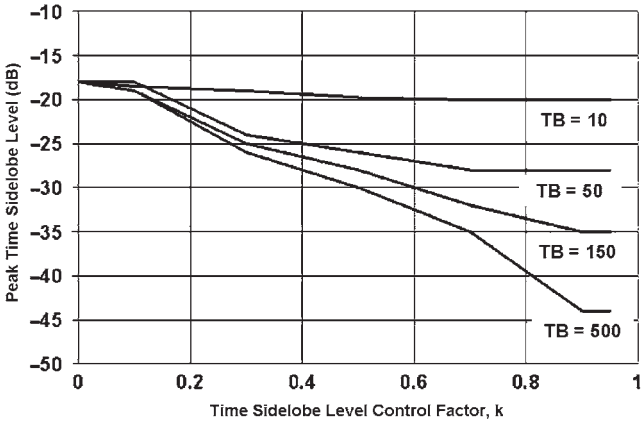


FIGURE 8.11 Peak time sidelobe level for a sine-based NLFM waveform as a function of k-factor (*Courtesy of Dr. Peter H. Stockmann, Lockheed Martin Maritime and Sensor Systems, Syracuse, NY*)

The frequency modulation-versus-time function for a tangent-based waveform is given as

$$f(t) = B\tan(2\beta t/T)/(2\tan\beta) \quad \text{for } -T/2 \le t \le T/2 \tag{8.16}$$

where T is the pulsewidth, B is the swept bandwidth, and β is defined as

$$\beta = \tan^{-1}(\alpha),\ 0 \le \alpha < \infty$$

where α is a time sidelobe level control factor.

When α is zero, the tangent-based NLFM waveform reduces to an LFM waveform. However, α cannot be made arbitrarily large because the compressed pulse tends to distort. Collins and Atkins[15] discuss an extension of the tangent-based NLFM for which the frequency modulation function is a weighted sum of tangent-based and linear frequency modulation terms.

Figure 8.12 shows the frequency modulation-versus-time functions for a sine-based NLFM waveform with $k = 0.6$, a tangent-based NLFM waveform with $\alpha = 2.5$, and an LFM waveform.

The sensitivity of a NLFM waveform to doppler shift can be seen in Figure 8.13, which shows the matched filter output for a sine-based NLFM waveform in the presence of doppler shift.

The ambiguity function of a NLFM sine-based waveform is shown in Figure 8.14. It can be noted that this ambiguity function is more thumbtack-like in nature than for an LFM waveform, indicating that this waveform is more doppler sensitive than the LFM waveform.

Table 8.3 provides a comparison of NLFM waveforms with weighted and unweighted LFM for different values of the target radial velocity in terms of peak and average time sidelobe levels and SNR loss. The NLFM waveform shows better

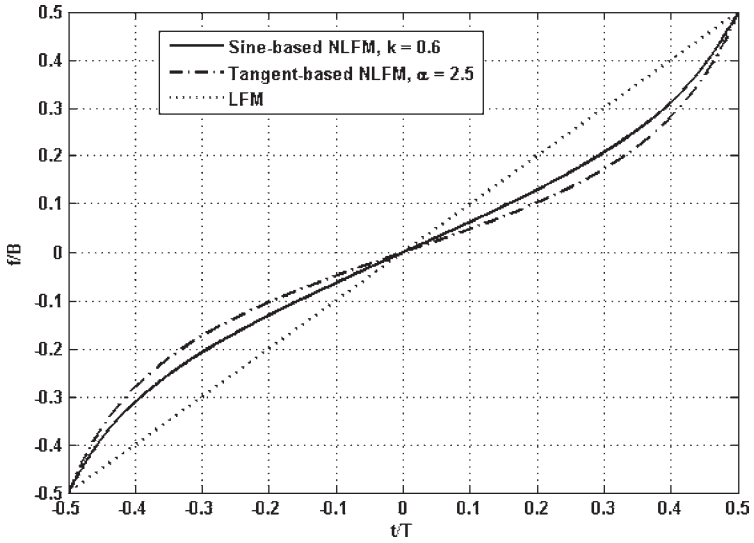


FIGURE 8.12 Frequency modulation-versus-time for sine-based NLFM, tangent-based NLFM, and LFM waveforms

performance in terms of SNR loss and peak time sidelobe level (TSL) than the LFM waveform. The TSL level does not degrade appreciably for the LFM waveform for higher radial velocities, demonstrating the higher doppler tolerance of LFM.

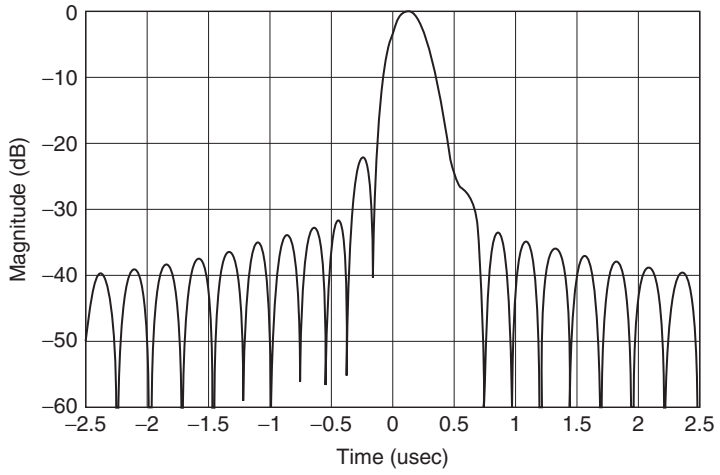


FIGURE 8.13 Matched filter output of S band, 44-μs pulsewidth, 5-MHz bandwidth NLFM sine-based waveform with 500 m/s radial velocity (*Courtesy of Edwin M. Waterschoot, Lockheed Martin Maritime and Sensor Systems, Syracuse, NY*)

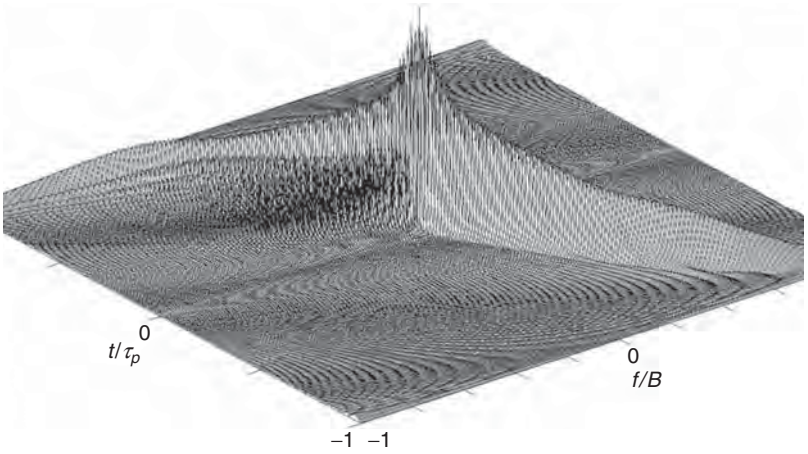


FIGURE 8.14 Ambiguity function of a sine-based symmetrical NLFM waveform

Phase-Coded Waveforms. In phase-coded waveforms, the pulse is subdivided into a number of subpulses each of duration $\delta = T/N$ where T is the pulsewidth and N is the number of subpulses. Phase-coded waveforms are characterized by the phase modulation applied to each subpulse.

Binary Phase Codes. A phase-coded waveform that employs two phases is called *binary*, or *biphase*, coding. A binary phase-coded waveform is constant in magnitude with two phase values, 0° or 180°. The binary code consists of a sequence of either 0s and 1s or +1s and −1s. The phase of the signal alternates

TABLE 8.3 Comparison of Linear FM and Nonlinear FM Waveform Performance*

Weighting	Target Radial Velocity (m/s)*	Peak TSL (dB)	Average TSL (dB)**	Filter Matching Loss (dB)
LFM unweighted	0	–13.32	–36.59	0
LFM unweighted	±300	–13.32	–36.56	0.024
LFM with –33 dB Taylor weighting	0	–32.43	–49.27	0.843
LFM with –33 dB Taylor weighting	±300	–32.25	–49.25	0.845
Sine-based NLFM with $k = 0.70$	0	–32.67	–48.97	0
Sine-based NLFM with $k = 0.70$	±300	–26.07	–47.99	0.038

* An S-band radar with 44-μs transmit pulsewidth and 5-MHz bandwidth was used in this comparison. The doppler shift expressed in Hz is $f_d = -(2/\lambda)V_r = -20V_r$ where V_r is the radial velocity expressed in m/s ($V_r > 0$ for an out-bound target).

** Average of TSL power ratio

between 0° and 180° in accordance with the sequence of elements, 0s and 1s or +1s and −1s, in the phase code, as shown in Figure 8.15. Because the frequency is not usually a multiple of the reciprocal of the subpulse width, the coded signal is generally discontinuous at the phase-reversal points This does not impact its time sidelobes, but does cause some increase in the spectrum sidelobe levels.[17]

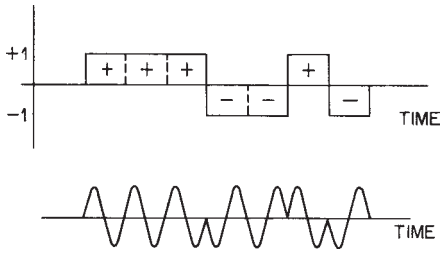


FIGURE 8.15 Binary phase-coded signal

Upon reception, the compressed pulse is obtained by matched-filter processing. The width of the compressed pulse at the half-amplitude point is nominally equal to the subpulse width. The range resolution is hence proportional to the time duration of one element of the code (one subpulse). The time-bandwidth product and pulse compression ratio are equal to the number of subpulses in the waveform; i.e., the number of elements in the code.

Optimal Binary Codes. Optimal binary codes are binary sequences whose peak sidelobe of the aperiodic autocorrelation function is the minimum possible for a given code length. Codes whose autocorrelation function, or zero-doppler response, exhibit low sidelobes are desirable for pulse compression radars. Responses due to moving targets will differ from the zero-doppler response. If the matched filter is based only on the zero-doppler response, an increase in the time sidelobes will result. Ultimately, if the doppler shift becomes very large, the matched filter response will degrade. This can be alleviated by utilizing a bank of matched filters, covering the expected range of doppler shifts. Because this is more computationally intensive than a single matched filter, older radar systems tend not to employ banks of filters. The increase in computational capacity of modern radar systems, however, can make this more attractive.

Barker Codes. A special class of binary codes is the Barker[18] codes. Barker codes are binary codes with peak time sidelobe levels equal to $-20\log(N)$, where N is the length of the code. The energy in the sidelobe region is minimum and uniformly distributed.[19] The Barker code is the only uniform phase code that reaches this level.[20] All the known binary Barker codes are listed in Table 8.4. Only binary Barker codes of lengths 2, 3, 4, 5, 7, 11, and 13 have been found.[21–24]

A pulse compression radar using Barker codes would be limited to a maximum time-bandwidth product of 13.[26] Figure 8.16 shows the autocorrelation function of

TABLE 8.4 Known Binary Barker Codes[25]

Length	Code
2	11, 10
3	110
4	1101, 1110
5	11101
7	1110010
11	11100010010
13	1111100110101

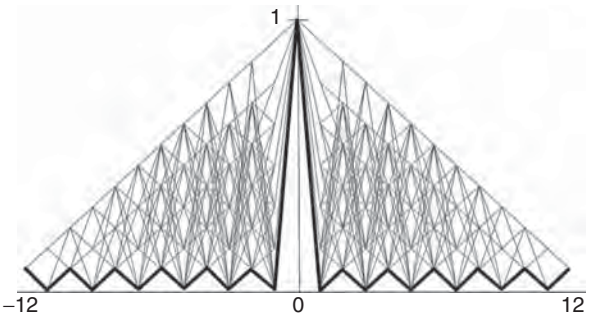


FIGURE 8.16 Superposition of the autocorrelation functions for all possible 13-bit code sequences with the Barker Code highlighted (dark), shown for zero doppler shift

a length 13 Barker code for zero doppler shift superimposed upon all possible autocorrelation functions of 13-bit binary sequences. It can be seen that the Barker code provides the lowest time sidelobe levels of all possible codes.

Allomorphic Forms. A binary code may be represented in any one of four allomorphic forms, all of which have the same correlation characteristics. These forms are the code itself, the inverted code (the code written in reverse order), the complemented code (1s changed to 0s and 0s to 1s), and the inverted complemented code. For symmetrical codes, the code and its inverse are identical.

Maximal-Length Sequences. Maximal-length sequences have a structure similar to random sequences and, therefore, possess desirable autocorrelation functions. They are often called *pseudorandom noise (PRN)* sequences. Historically, these sequences were generated using n stages of shift registers with selected output taps used for feedback (see Figure 8.17). When the feedback connections are properly chosen, the output is a sequence of maximal length, which is the maximum length of a sequence of 1s and 0s that can be formed before the sequence is repeated. The length of the maximal sequence is $N = 2^n - 1$, where n is the number of stages in the shift-register generator.

The feedback connections that provide the maximal-length sequences may be determined from a study of primitive and irreducible polynomials. An extensive list of these polynomials is given by Peterson and Weldon.[27]

Although maximal-length sequences have some desirable autocorrelation characteristics, a maximum length sequence does not guarantee lowest time sidelobes when compared to other binary codes. An example of this is provided for a 15-bit sequence. Figure 8.18*a* is a histogram of the peak time sidelobe level for the autocorrelation of every possible combination of a 15-bit code. Figure 8.18*b* is the same but for only maximal length sequences of length 15 code (a subset of Figure 8.18*a*). Figure 8.18*a* shows a lowest time sidelobe level of –17.5 dB. The lowest sidelobe for the maximal length sequence is seen from Figure 8.18*b* to be only –14 dB.

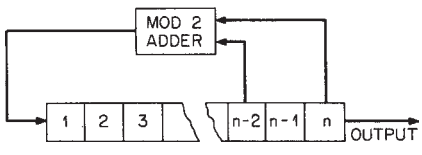


FIGURE 8.17 Shift-register generator

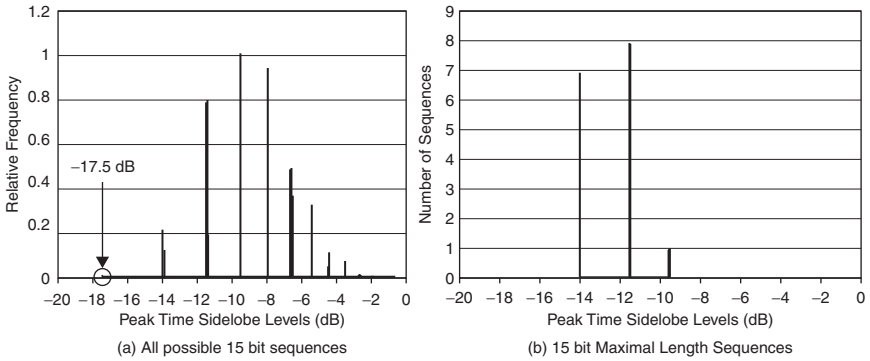


FIGURE 8.18 Histogram of peak time sidelobe levels for 15-bit sequences: (*a*) all possible 15-bit sequences and (*b*) 15-bit maximal length sequences

Minimum Peak Sidelobe Codes. Binary codes that provide minimum peak time sidelobe levels but exceed the time sidelobe levels achieved by Barker codes (–20 log (*N*)) are termed minimum peak sidelobe codes.[29] These codes are usually found using computer search techniques. Skolnik[28] and Levanon and Mozeson[29] provide these codes for various sequence lengths, along with the resulting time sidelobe levels.

Complementary Sequences. Complementary sequences consist of two sequences of the same length N whose aperiodic autocorrelation functions have sidelobes equal in magnitude but opposite in sign. The sum of the two autocorrelation functions has a peak of $2N$ and a sidelobe level of 0. In a practical application, the two sequences must be separated in time, frequency, or polarization, which results in decorrelation of radar returns so that complete sidelobe cancellation may not occur. Hence, they have not been widely used in pulse compression radars.

Polyphase Codes. Waveforms consisting of more than two phases may also be used. Polyphase codes can be considered as complex sequences whose elements have a magnitude of one, but with variable phase.[30] The phases of the subpulses alternate among multiple values rather than just the 0° and 180° of binary phase codes. These codes tend to be discrete approximations to LFM waveforms, and hence possess similar ambiguity functions and doppler shift characteristics. The autocorrelation functions are similar, with a peak to sidelobe ratio of about $\sqrt{N}$.

Frank Codes. The Frank code corresponds to a stepped-phase approximation of an LFM waveform.[31] Here, the pulse is broken up into M groups, each of which is further broken up into M subpulses. Hence, the total length of the Frank code, is M^2, with a corresponding compression ratio of M^2. The Frank polyphase codes[32] derive the sequence of phases for the subpulses by using a matrix technique as follows:

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 2 & \cdots & (M-1) \\ 0 & 2 & 4 & \cdots & 2(M-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & (M-1) & 2(M-1) & \cdots & (M-1)^2 \end{bmatrix} \tag{8.17}$$

The matrix elements represent the multiplying coefficients of a basic phase shift $2\pi/M$, where M is an integer. The phase shift corresponding to the element m,n of the matrix can be written as

$$\phi_{m,n} = \frac{2\pi}{M}(m-1)(n-1),\ m = 1,\ldots,M,\ n = 1,\ldots,M \tag{8.18}$$

An example of a Frank Code matrix for $M = 4$ is given here:

$$\frac{\pi}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 6 \\ 0 & 3 & 6 & 9 \end{bmatrix} = \frac{\pi}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 0 & 2 \\ 0 & 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 90^\circ & 180^\circ & 270^\circ \\ 0 & 180^\circ & 0 & 180^\circ \\ 0 & 270^\circ & 180^\circ & 90^\circ \end{bmatrix}$$

Concatenating the rows of this matrix yields the phase for each of the 16 subpulses. Figure 8.19 shows the phase modulation characteristic of the Frank Code for the above example. Note how the phase step between subpulses increases between subpulse groups with a length equal to four. This characteristic can be regarded as a stepped-phase approximation to quadratic phase modulation.

As M increases, the peak-sidelobe–voltage ratio approaches $(\pi M)^{-1}$. This corresponds to approximately a 10-dB improvement over pseudorandom sequences of similar length. The ambiguity function grossly resembles the knife-edge (ridge) characteristic associated with LFM waveforms, as contrasted with the thumbtack characteristic of pseudorandom sequences (Figure 8.20) However, for small ratios of doppler shift to waveform bandwidth, a good doppler response can be obtained for reasonable target velocities.

Lewis and Kretschmer Codes (P1, P2, P3, P4). Lewis and Kretschmer have studied the P1, P2, P3, and P4 polyphase codes.[33,35] These codes are step approximations to the LFM pulse compression waveforms,[34] have low-range sidelobes, and have the

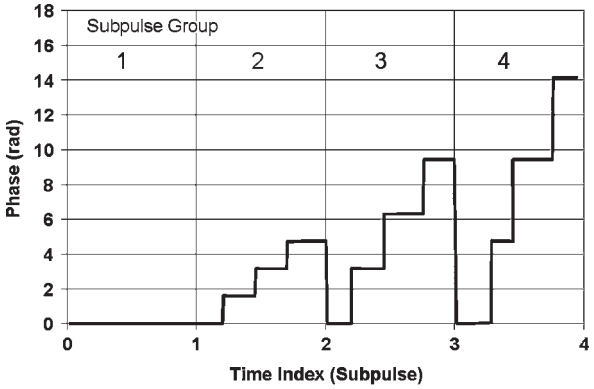


FIGURE 8.19 Phase versus time relationship for Frank code of length 16 ($M = 4$)

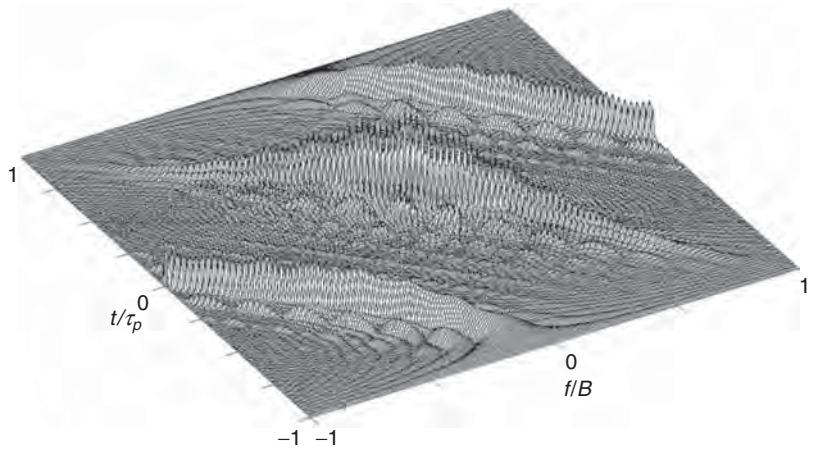


FIGURE 8.20 Ambiguity function of a Frank code of length 64 ($M = 8$)

doppler tolerance of the LFM codes. The P1 and P2 codes are modified versions of the Frank code with the DC frequency term at the center of the pulse instead of at the beginning. They are more tolerant of receiver band-limiting prior to pulse compression encountered in digital radar systems. The P1 codes contains M^2 elements as does the Frank code, but the relationship of the ith element to the jth group is expressed as[35]

$$\phi_{i,j} = -(\pi/M)[M-(2j-1)][(j-1)M+(i-1)] \tag{8.19}$$

where i and j are integers ranging from 1 to M.

P2 codes are similar, but the phase is symmetric with the following characteristic:

$$\phi_{i,j} = \{(\pi/2)[(M-1)/M]-(\pi/M)(i-j)\}[M+1-2j] \tag{8.20}$$

The P3 and P4 codes are derived by essentially converting an LFM waveform to baseband.[35] These tend to be more doppler tolerant than the Frank, P1, or P2 codes, and are also more tolerant of precompression bandwidth limitations that appear in radar systems. The phase of the P3 code is given as

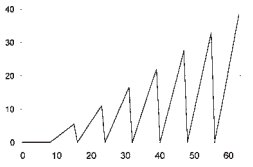
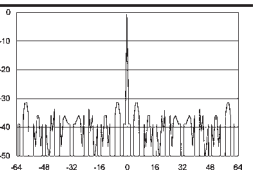
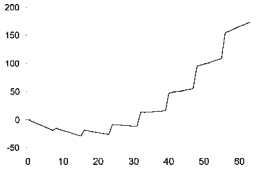
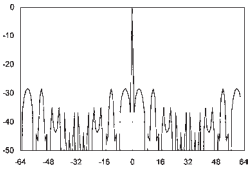
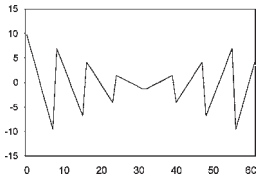
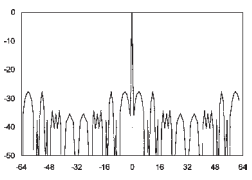
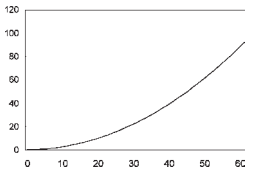
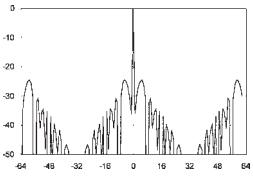
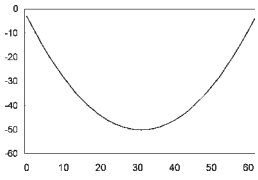
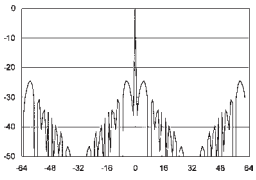
$$\varphi_n = \frac{\pi}{N}n^2 \qquad n = 0,\ldots,N-1$$

The P4 code phase relationship is similar:

$$\phi_n = \frac{\pi n^2}{N} - \pi k \qquad 0 \le n \le N$$

Table 8.5 summarizes the phase and autocorrelation characteristics of the Frank code and the Lewis and Kretschmer P1 through P4 polyphase codes.

TABLE 8.5 Summary of Phase and Autocorrelation Characteristics of Frank and Lewis and Kretschmer Polyphase Codes

Polyphase Code	Phase	Phase vs. Time Characteristic ($N = 64$ Example)	Autocorrelation (dB) ($N = 64$ Example)
Frank	$\frac{2\pi}{N}(i-1)(j-1)$ $i = 1,2,3,\ldots N$ $j = 1,2,3\ldots N$		
P1	$\frac{-\pi}{M}[M-(2j-1)]$ $\bullet\,[(j-1)M+(i-1)]$ for ith element in the jth group		
P2	$\{(\pi/2)[(M-1)/M]$ $-(\pi/M)(i-j)\}$ $\bullet\,[M+1-2j]$ for ith element in the jth group		
P3	$\frac{\pi}{N}n^2$ $n = 0,\ldots. N-1$		
P4	$\frac{\pi n^2}{N}-\pi k$ $0 \le n \le N$		

P(n,k) Polyphase Codes. Whereas the previously discussed polyphase codes are derived from LFM waveforms, $P(n, k)$ codes are derived from step approximations of the phase characteristic of the weighting function of NLFM waveforms.[20] The weighting function is given by

$$W(f) = k + (1-k)\cos^n\left(\frac{\pi f}{B}\right) \tag{8.21}$$

where k and n are parameters of the weighting function, B is the swept bandwidth of the waveform, and $-B/2 \le f \le B/2$. This is a $\cos^n$ weighting on a pedestal of height k (Figure 8.21). Hamming weighting is achieved for $n = 2$ and $k = 0.08$.

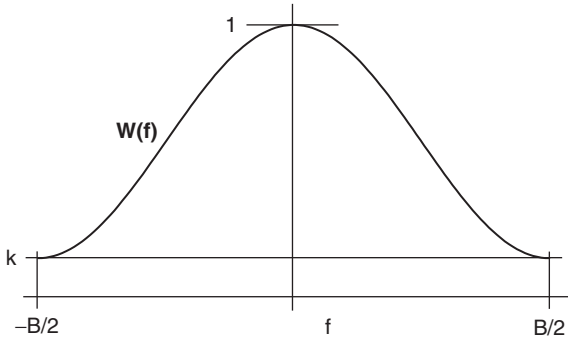


FIGURE 8.21 $\cos^n$ on pedestal weighting function (shown for $n = 2$)

For the case where $n = 2$, the weighting function can be integrated to obtain the following relationship between time and frequency[31]

$$\frac{t}{T} = \frac{f}{B} + a\sin(2\pi f/B) \quad \text{where } a = (1-k)/2(1+k) \tag{8.22}$$

which is similar to the sine-based NLFM discussed earlier. This particular code is called Phase from Nonlinear Frequency (PNL)[31] and its autocorrelation function is shown in Figure 8.22 for a 100-μs pulsewidth, 1-MHz bandwidth waveform with $a = 0.7$ and $f_d = 0$. The time sidelobe levels are seen to be below –32 dB.

The ambiguity function is similar to that provided in the discussion of NLFM, which is expanded in Figure 8.23 to show in more detail the impact of doppler shift on the pulse compressed waveform for practical values of doppler shifts.

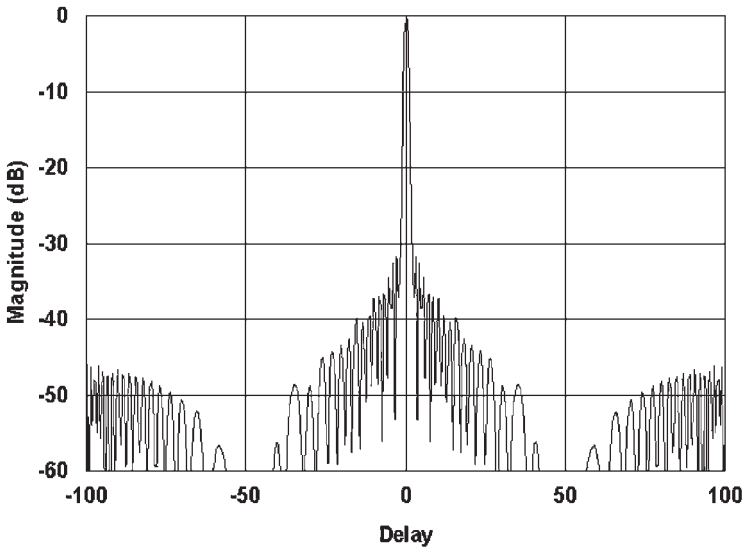


FIGURE 8.22 100-μs PNL pulse autocorrelation function for $TB = 100$, $a = 0.7$, and $f_d = 0$

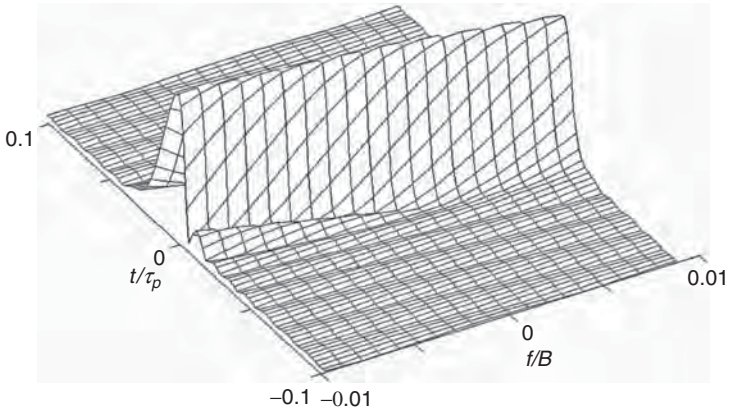


FIGURE 8.23 Expanded view of PNL ambiguity diagram for 100-μs pulse, $a = 0.7$, and $B = 1$ MHz

As the doppler shift moves away from zero, the peak decreases and the close-in time sidelobe levels on one side or the other begin to increase. Note that an *f*/*B* ratio of 0.01 corresponds to a doppler shift associated with approximately a Mach 1 target at a S band carrier frequency.

In general for $P(n, k)$ waveforms, the integral of the weighting function provides the relationship between time and frequency modulation as shown in Eq. 8.23:

$$\frac{t}{T} = \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi f}{B}} [k + (1-k)\cos^n(x)]dx \tag{8.23}$$

Since frequency modulation is proportional to the time derivative of phase, phase is obtained by integrating the frequency with respect to time. The expression for frequency, however, is not straightforward, and is usually obtained through numerical evaluation.[31]

Quadriphase Codes. Quadriphase codes are an example of a phase-coded waveform without phase discontinuities. Quadriphase codes[36,37] are based on the use of subpulses with a half-cosine shape and phase changes between adjacent subpulses of multiples of ±90°. The cosine weighting provides faster spectrum roll-off, lower filter matching loss, and smaller range sampling loss when compared to rectangular subpulse phase-coded waveforms (Table 8.6).

TABLE 8.6 Quadriphase Waveform Performance Summary[36]

	Quadriphase Code	Rectangular Subpulse Code
Radiated Spectrum 40-dB Width	$5/\delta$	$64/\delta$
Falloff (δ= subpulse duration)	12 dB/Octave	6 dB/Octave
Range Sampling Loss	0.8 dB	2.3 dB
Filter Matching Loss	0.1 dB	0.5 dB

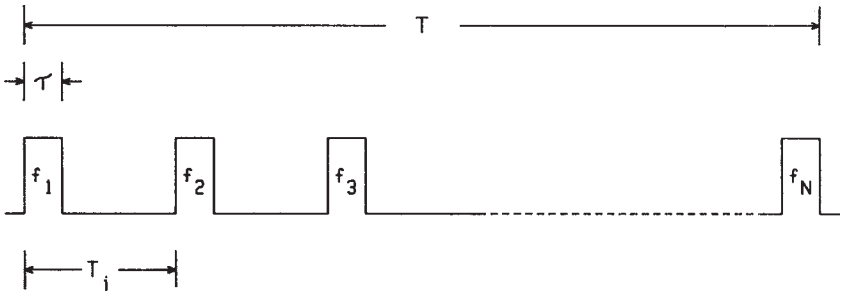


FIGURE 8.24 Time-frequency-coded waveform

Time-Frequency-Coded Waveforms. A time-frequency-coded waveform (Figure 8.24) consists of a train of N pulses with each pulse at a different frequency.[38] Generally, the frequencies are equally spaced, and the pulses are of the same amplitude. The ambiguity function for a periodic waveform of this type consists of a central spike plus multiple spikes or ridges displaced in time and frequency. Although it is unachievable in practice, the objective is to create a high-resolution, thumbtack-like central spike with a clear area around it. Measurement is then performed on the high-resolution central spike. The range resolution or compressed-pulse width is determined by the total bandwidth of all the pulses, and the doppler resolution is determined by the reciprocal of the waveform duration T. For example, a typical waveform in this class has N contiguous pulses of width τ whose spectra of width $1/\tau$ are placed side by side in frequency to eliminate gaps in the composite spectrum. Since the waveform bandwidth is now N/τ, the nominal compressed-pulse width is τ/N. These relationships are summarized in Table 8.7.

Shaping of the high-resolution central spike area as well as the gross structure of the ambiguity surface can be accomplished by variations of the basic waveform parameters, such as amplitude weighting of the pulse train, staggering of the pulse repetition interval, and frequency or phase coding of the individual pulses.[39]

Costas Codes. Costas codes are a class of frequency-coded waveforms that have near ideal range and doppler sidelobe behavior.[40,41] In other words, their ambiguity function approaches the ideal thumbtack, providing both doppler and range information (Figure 8.25). All sidelobes, except for a few near the origin, have an amplitude of $1/M$. A few sidelobes close to the origin are about twice as large, or about $2/M$, which is characteristic of Costas codes. The compression ratio of a Costas code is about M.[42]

The Costas code is a burst of M contiguous uncoded pulse waveforms, each with a different frequency selected from a finite set of M equally spaced frequencies[43] that are

TABLE 8.7 *N* Pulses Contiguous in Time and Frequency

Waveform duration, T	$N\tau$
Waveform bandwidth, B	N/τ
Time-bandwidth product, TB	N^2
Compressed pulse width, $1/B$	$\tau/N = \tau/N^2$
Doppler resolution, $1/T$	$1/N\tau$

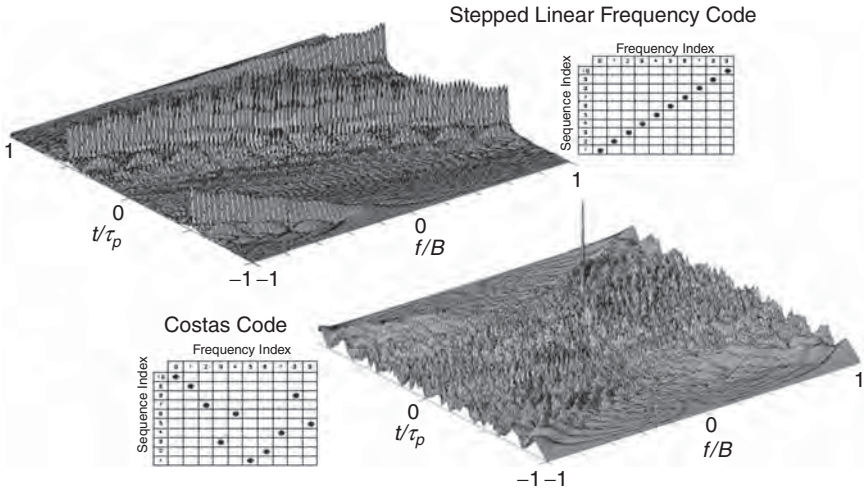


FIGURE 8.25 Comparison of ambiguity functions for $N = 10$ stepped linear and Costas sequence showing the impact of frequency order[40]

processed coherently. The order in which the frequencies are generated greatly influences the nature of the ambiguity function of the burst. If the frequencies are monotonically increasing or decreasing, the waveform is simply a stepped approximation to an LFM, which has a ridge in its ambiguity function (Figure 8.25). In order to approach a thumbtack-like ambiguity function, the order of the frequencies needs to be more random in nature. The order of frequencies is the code, and it is generated via a special class of $M \times M$ Costas arrays. Costas[41,44] suggested a technique for selecting the order of these frequencies to provide more controlled range and doppler sidelobes. An example of a Costas code of length 10 is shown in Figure 8.25, as it compares to the stepped LFM. Tables showing the sequence order for each waveform are also provided.

8.3 FACTORS AFFECTING CHOICE OF PULSE COMPRESSION SYSTEMS

The choice of a pulse compression system involves the selection of the type of waveform and the method of generation and processing. Methods of generating and processing pulse compression waveforms are discussed in the section on pulse compression implementation in this chapter. Discussions here will concentrate on the waveform itself. The primary factors influencing the selection of a particular waveform are usually the radar requirements of doppler tolerance and time sidelobe levels.

Table 8.8 summarizes these factors for three FM types: LFM, NLFM, and phase-coded waveforms. The systems are compared on the assumption that information is extracted by processing a single waveform as opposed to multiple-pulse processing. The symbols B and T denote the bandwidth and the pulsewidth of the waveform, respectively.

In cases where an insufficient doppler shift occurs, such as with a stationary or tangential target, range resolution is the chief means for seeing a target in clutter.

TABLE 8.8 Comparisons of Performance Characteristics for LFM, NLFM, and Phase-Coded Waveforms

Factor	Linear FM	Nonlinear FM	Binary Phase Coded	Polyphase Coded
Doppler tolerance	Supports doppler shifts up to $\pm B/10$. Time shift of f_dT/B is introduced by range-doppler coupling. Time sidelobe performance remains excellent for large doppler shifts.	Adequate insensitivity to doppler to allow use generally up to Mach 1. Time shift of f_dT/B is introduced by range-doppler coupling for a nonsymmetrical NLFM waveform. Common, therefore, in ATC radars. Multiple tuned pulse compressors required for high-speed targets.	Higher sensitivity to doppler shift. Time sidelobes increase while mainlobe response decreases for higher doppler (characteristic of a thumbtack-like ambiguity function). Used, therefore, for low-speed target applications and with small TB products.	Highest sensitivity to doppler shift. Time sidelobes increase while mainlobe response decreases for higher doppler (characteristic of a thumbtack-like ambiguity function). Used, therefore, for low-speed target applications and with small TB products. Longer phase-coded waveforms are more sensitive to doppler shifts than the shorter ones.
Time sidelobe level	Adequate weighting, high TB product, and low amplitude and phase errors are necessary to achieve good time sidelobes.	For nonsymmetrical NLFM, excellent time sidelobes if there is adequate NLFM phase coding, a high TB product, and sufficiently low amplitude and phase errors. Increasing NLFM phase code weighting introduces increased radial velocity sensitivity.	Good time sidelobes that are determined by coding.	Better time sidelobes than binary phase-coded waveforms.
General	Often used for high-speed target capability ($\gg$ Mach 1). Extremely wide bandwidths achievable.	Use is generally restricted to applications where primary target radial velocities < Mach 1. Multiple tuned matched filters are generally not computationally practical.	Generally found in low doppler shift applications.	Generally found in low doppler shift applications.

Clutter rejection with pulse compression waveforms is due to the greater range resolution achievable over uncoded waveforms. Because the range resolution is proportional to the reciprocal of the bandwidth, wider-bandwidth pulse compression waveforms can offer greater clutter rejection.

8.4 PULSE COMPRESSION IMPLEMENTATION AND RADAR SYSTEM EXAMPLES

This section describes the generation and processing of pulse compression waveforms and provides examples of radar systems that utilize these processing techniques. Major advances are continually being made in the devices and techniques used in pulse compression radars. Significant advances are evident in the digital and SAW techniques that allow the implementation of a variety of pulse compression waveform types. The digital approach has blossomed because of the manifold increase in computational speed and also because of the size reduction and the speed increase of the memory units. SAW technology has expanded because of the invention of the interdigital transducer,[45] which provides efficient transformation of an electrical signal into acoustic energy and vice versa.

Digital Waveform Generation. Figure 8.26 shows a digital approach[46] for generating the radar waveform. The phase control element supplies digital samples of the in-phase component I and the quadrature component Q, which are converted to their analog equivalents. These phase samples may define the baseband components of the desired waveform, or they may define the waveform components on a low-frequency carrier. If the waveform is on a carrier, the balanced modulator is not required, and the filtered components would be added directly. The sample-and-hold circuit removes the transients due to the nonzero transition time of the digital-to-analog (D/A) converter. The low-pass filter smoothes (or interpolates) the analog signal components between waveform samples to provide the equivalent of a much higher waveform-sampling rate. The $I(t)$ component modulates a 0° carrier signal, and the $Q(t)$ component modulates a 90° phase-shifted carrier signal. The desired waveform is the sum of the 0°-modulated carrier and the 90°-modulated carrier. As mentioned earlier, when the digital phase samples include the carrier components, the I and Q components are centered on this carrier frequency and the low-pass filter can be replaced with a bandpass filter centered on the IF carrier.

When a linear-FM waveform is desired, the phase samples follow a quadratic pattern and can be generated by two cascaded digital integrators. The input digital command to the first integrator defines this quadratic phase function. The digital command to the second integrator is the output of the first integrator plus the desired carrier frequency. This carrier may be defined by the initial value of the first integrator. The desired initial phase of the waveform is the initial value of the second integrator or else may be added to the second-integrator output.

With advances in digital technology, it has become possible and practical to generate waveforms directly at IF or RF carrier frequencies on a single integrated circuit chip. This technique is called *Direct Digital Synthesis*, or *DDS*, and involves generating waveforms at

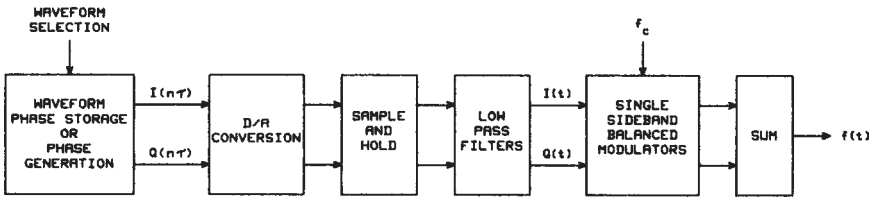


FIGURE 8.26 Digital waveform generation block diagram

high sampling rates and filtering the output. These devices generate the waveform by accumulating phase information, which is then used to look up values of the waveform (usually a sine wave). This is converted to an analog signal with a digital-to-analog converter (DAC or D/A converter) and filtered. A variety of waveform types (e.g., LFM, NLFM, and CW waveforms) can be generated in this way by using the appropriate phase modulation characteristic. As an example, the Analog Devices AD9858 Direct Digital Synthesizer[47] uses a 10-bit DAC operating at up to a 1-GHz internal clock speed (DAC update rate).

Digital Pulse Compression.[48–50] Digital pulse compression techniques are routinely used for matched filtering of radar waveforms. The matched filter may be implemented by using a digital convolution for any waveform or else by use of stretch processing for a linear-FM waveform.

Digital pulse compression has distinct features that determine its acceptability for a particular radar application. Digital matched filtering usually requires multiple overlapped processing units for extended range coverage. The advantages of the digital approach are that long-duration waveforms present no problem, the results are extremely stable under a wide variety of operating conditions, and the same implementation could be used to handle multiple waveform types.

Analog product detectors used to extract I and Q baseband components have been replaced in many systems by digital down-conversion techniques. In this approach, the complex envelope sequence is evaluated by digital signal processing of A/D converter samples at the final IF output of the receiver, rather than by separate A/D conversion of baseband analog I and Q components.[51–53] Digital down-conversion is advantageous because performance is not limited by amplitude and phase imbalances that exist in analog product-detection hardware.

Figure 8.27 illustrates two digital signal-processing approaches to providing the matched filter for a pulse compression waveform. In both cases, the input signal is the complex envelope sequence as formed using either digital down-conversion or analog

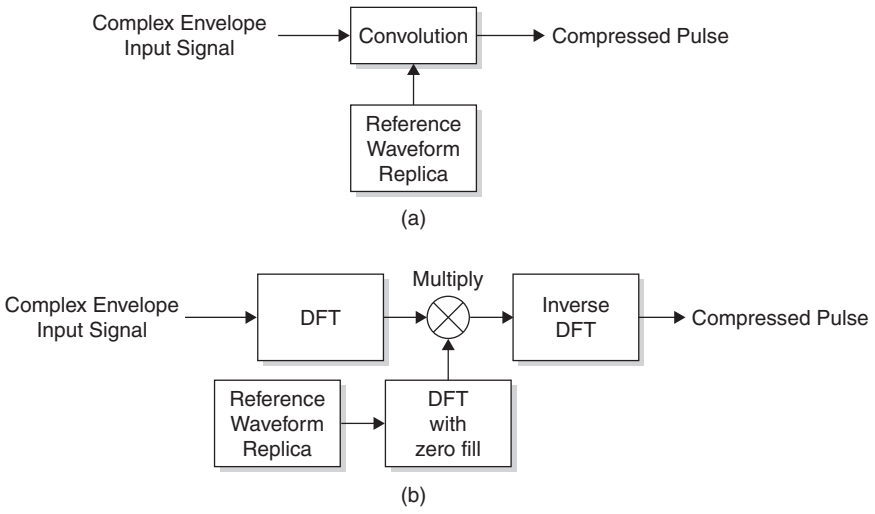


FIGURE 8.27 (*a*) Time-domain digital pulse compression processor and (*b*) frequency-domain digital pulse compression processor

product detection followed by A/D conversion in each baseband channel. Figure 8.27*a* shows a digital implementation of a time-domain convolution processor that will provide matched-filter performance for any radar waveform. In this case, discrete-time convolution is done in the time domain by convolution of the complex envelope input sequence following digital down-conversion with the matched filter impulse response sequence. Because time-domain convolution can be computationally intensive, a more economical approach from a computational standpoint is shown in Figure 8.27*b*, in which frequency-domain processing is used to implement the convolution.

The frequency-domain digital pulse compression processor operates on the principle that the discrete Fourier transform (DFT) of the time convolution of two sequences is equal to the product of the discrete Fourier transforms of each of the sequences. If M range samples are to be provided by one processor, the length of the DFT must exceed M plus the number of samples in the reference waveform minus one to achieve an aperiodic convolution. These added M samples are filled with zeros in the reference waveform DFT. For extended range coverage, repeated processing operations are required with range delays of M samples between adjacent operations using the overlap-save convolution technique.[49,54] This processor can be used with any waveform, and the reference waveform can be offset in doppler frequency to achieve a matched filter at this doppler frequency.

Pulse Compression Radar Examples. There are many radars under development or deployed that utilize some of the pulse compression waveforms previously discussed. Advances in digital signal-processing technology have enabled a wider variety of waveform implementations. For example, radar systems are no longer limited to the LFM waveform; instead, radar system capabilities can be extended to take advantage of the more complex processing associated with the nonlinear FM waveform.

AN/TPS-59 and AN/FPS-117 Surveillance Radars.[55] The AN/TPS-59 and AN/FPS-117 are a family of L band, long-range surveillance radars that employ LFM waveforms. The antenna is mechanically rotated in azimuth, and electronic pencil-beam scanning is performed in elevation. The transmission utilizes two time-sequenced LFM pulses of different frequencies in order to create Swerling Case 2 target statistics. Both radars employ frequency-domain digital pulse compression processing.

Air Surveillance and Precision Approach Radar System. The Air Surveillance and Precision Approach Radar System (ASPARCS) is intended to provide the next-generation air traffic control (ATC) radar, as part of the Multi-Mission Surveillance Radar (MMSR) family of ATC radars built by Lockheed Martin Co. Nonlinear FM waveforms are used because the targets of interest have relatively low doppler shifts (less than Mach 1). Like the AN/FPS-117 radar, this system implements frequency-domain digital pulse compression processing.

Multi-Mission Radar. The Multi-Mission Radar (MMR) is designed to detect and track mortars, artillery, and rockets. This radar uses a nonlinear FM sine-based waveform. Digital frequency-domain pulse compression processing is performed.

ASR-12 Next-Generation Solid-State Air Traffic Control Radar.[56] The ASR-12 terminal airport surveillance radar transmits a 55-μs pulse with peak power of 21 kW to provide a single-pulse transmit energy of 1.16 J. Nonlinear frequency modulation is used with a pulse compression ratio of 55 to achieve range-resolution equivalent to an uncoded 1-μs pulse. The filter matching loss is less than 0.6 dB and typical time

sidelobe levels measured on production hardware are –58 dB. Digital pulse compression is used. An uncoded 1.1-μs pulse is used to provide coverage for targets within the range interval from 0.5 to 5.5 nmi.

Stretch Pulse Compression.[57–60,62] Stretch pulse compression is a technique for performing LFM pulse compression of wideband waveforms using a signal processor with bandwidth that is much smaller than the waveform bandwidth, without loss of signal-to-noise ratio or range resolution. Stretch pulse compression is used for a single target or for multiple targets that are located within a relatively small range window centered at a selected range.

Figure 8.28 shows a block diagram of a stretch pulse compression system. The LFM waveform has a swept bandwidth B, pulsewidth T, and LFM slope α. The reference waveform is generated with time delay τ_R, swept bandwidth B_R, pulsewidth T_R, and LFM slope α_R. The reference waveform time delay is typically derived by range tracking of a selected target within the range window. The correlation mixer[62,63] (CM) in Figure 8.29 performs a bandpass multiplication of the received signal by the output of the reference waveform generator. The lower sideband at the CM output is selected by a bandpass filter (BPF).

Spectrum analysis is performed when the LFM slopes of the transmit and reference waveforms are equal ($\alpha = \alpha_R$). Reduced-bandwidth pulse compression processing is performed if the reference waveform LFM slope is less than the transmit waveform LFM slope ($\alpha_R < \alpha$). In both cases, the required processing bandwidth B_p is much smaller than the waveform bandwidth.

Figure 8.29 shows the principle of stretch pulse compression for the case where the LFM slopes of the transmit and reference waveforms are equal. The instantaneous frequency is plotted as a function of time at three points in the stretch pulse compression system block diagram: (1) correlation mixer input, (2) correlation mixer LO (reference waveform generator output), and (3) correlation mixer output (output of bandpass filter). Three LFM target signals are shown at the correlation mixer input: target 1 is at zero time offset relative to the reference waveform; target 2 is earlier in time than the reference waveform; and target 3 is later in time. In each case, the LFM slope for the target signals is B/T. The reference waveform applied to the LO port of the CM has LFM slope equal to $B_R/T_R = B/T$.

The instantaneous frequency at the correlation mixer output is the difference between the instantaneous frequencies at the CM input and LO ports. As a result, the CM output signals for the three target signals are uncoded pulses (pulsed CW signals) with frequency offset from the mixer IF output f_{IF} given by

$$\delta f = -\left(\frac{B}{T}\right) t_d \tag{8.24}$$

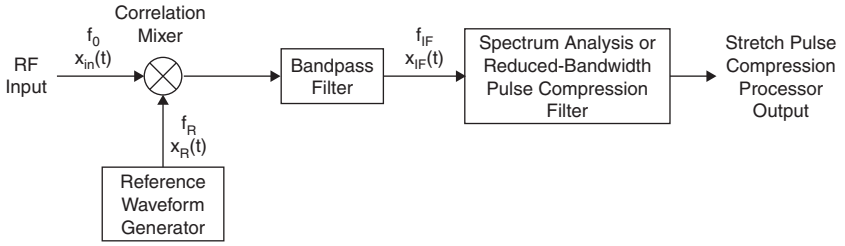


FIGURE 8.28 Stretch pulse compression system block diagram

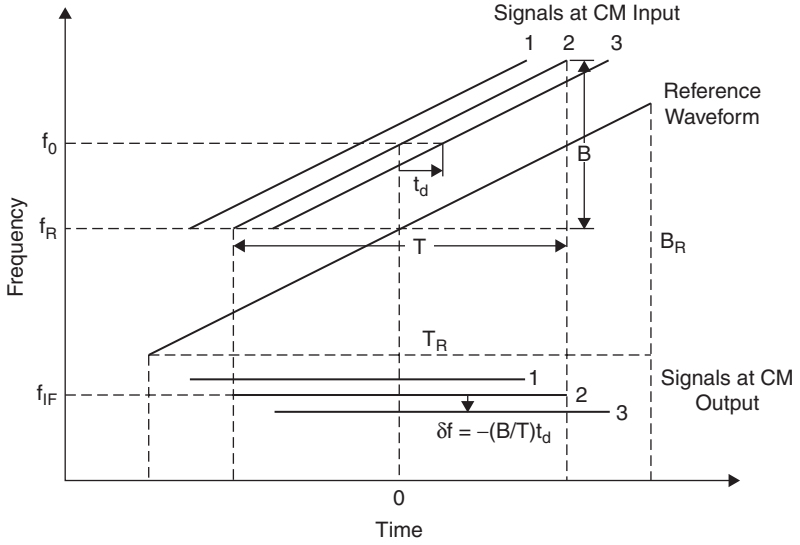


FIGURE 8.29 Correlation mixer signals in stretch pulse compression (*after Roth et al.*[61])

where t_d is the time delay of the midpoint of the signal measured relative to the midpoint of the reference waveform. For the case shown, where the RF carrier frequency is above the carrier frequency of the reference waveform, a positive time delay results in a negative frequency offset. The signals at the correlation mixer output are then resolved in the frequency domain by spectral analysis processing.

A typical implementation for the spectral analysis processing includes a second frequency conversion following the CM to a final intermediate frequency (IF), antialiasing filtering, direct sampling at the final IF using an analog-to-digital converter (ADC), digital down conversion (DDC) to a complex envelope sequence, time-domain weighting, and spectral analysis using an FFT padded with zeros.[64] Previous implementations used analog product detectors to extract I and Q baseband signals, with separate ADCs in the I and Q baseband channels.

Correlation Mixer Output Signal Analysis. The received signal at the CM input port from a point target is

$$x_{\text{in}}(t) = A\,\text{rect}\left(\frac{t-\tau}{\text{T}}\right)\cos[2\pi(f_0 + f_d)(t-\tau) + \pi\alpha(t-\tau)^2] \tag{8.25}$$

where A is the amplitude, T is the transmit pulsewidth, f_0 is the carrier frequency, f_d is the doppler frequency, τ is the signal time delay, and α is the LFM slope for the transmit waveform. The reference waveform applied to the LO port is

$$x_R(t) = 2\,\text{rect}\left(\frac{t-\tau_R}{T_R}\right)\cos[2\pi f_R(t-\tau_R) + \pi\alpha_R(t-\tau_R)^2] \tag{8.26}$$

where T_R is the pulsewidth, f_R is the carrier frequency, τ_R is the reference waveform time delay, and α_R is the LFM slope for the reference waveform ($\alpha_R \leq \alpha$).

The correlation mixer acts as a bandpass multiplier with output $x_{\text{in}}(t)x_R(t)$. The IF output of the correlation mixer is evaluated using the identity

$$2\cos x \cos y = \cos(x + y) + \cos (x - y)$$

where the first term on the right-hand side of the equation corresponds to the upper sideband and the second to the lower sideband at the mixer output. The upper sideband is rejected by the bandpass filter to yield

$$x_{\text{IF}}(t) = A\,\text{rect}\left(\frac{t-\tau}{T}\right)\text{rect}\left(\frac{t-\tau_R}{T_R}\right)$$
$$\cdot\cos[2\pi f_{\text{IF}}(t-\tau) + 2\pi f_d(t-\tau) + 2\pi\alpha_R(\tau_R-\tau)(t-\tau) + \pi(\alpha-\alpha_R)(t-\tau)^2 + \phi] \tag{8.27}$$

where $f_{\text{IF}} = f_0 - f_R$ is the IF carrier frequency ($f_0 > f_R$ is assumed) and the carrier phase shift is

$$\phi = -2\pi f_R(\tau - \tau_R) - 2\pi\alpha_R(\tau - \tau_R)^2$$

The IF signal is an LFM waveform with reduced slope $\alpha - \alpha_R$ (the factor that multiplies the quadratic term in the argument of the cosine) and a frequency offset relative to the IF carrier frequency f_{IF} given by

$$\delta f = f_d + \alpha_R(\tau_R - \tau) \tag{8.28}$$

The duration of the reference waveform is required to exceed the transmit pulsewidth to avoid a loss in SNR caused by target echoes that are not contained within the reference waveform.

Equal Transmit and Reference Waveform LFM Slopes. For the case where the transmit and reference waveform LFM slopes are equal ($\alpha = \alpha_R$), the IF signal is an uncoded pulse with frequency offset given by

$$\delta f = f_d + \alpha(\tau_R - \tau) \tag{8.29}$$

The frequency offset is measured using spectrum analysis and converted to target time delay and range relative to the reference waveform by

$$\Delta\tau = \tau - \tau_R = -\frac{\delta f}{\alpha}$$
$$\Delta r = (R - R_0) = \frac{c}{2}\Delta\tau \tag{8.30}$$

where $R_0 = c\tau_R/2$ is the range corresponding to the time delay of the reference waveform.

Kellog[65] describes additional considerations for application of time-domain weighting in stretch processing and provides details on compensation techniques for hardware errors. The effect of time mismatch between the signal and the weighting function is analyzed by Temes.[66]

Unequal Transmit and Reference Waveform Slopes. A stretch processor with unequal frequency-slope waveforms requires pulse compression of the residual linear FM at the output of the correlation mixer. A linear FM signal with a slope of $\alpha_{\text{in}} - \alpha_R$ occurs at the target range, which is offset in frequency from the IF carrier frequency by $\alpha_R(\tau_R - \hat{\tau})$. With the range-doppler coupling of the LFM waveform, the apparent time delay of this target will be

$$\tau_{\text{app}} = -\alpha_R(\tau_R - \tau)/(\alpha - \alpha_R) \tag{8.31}$$

This result can be interpreted as yielding a time-expansion factor of $\alpha_R/(\alpha - \alpha_R)$ for the compressed pulse. As for the case of equal LFM slopes, the range window width depends on the achievable processing bandwidth.

Stretch Processing Range Resolution Width. The 6-dB frequency resolution width for spectral analysis using a rectangular window of time duration equal to the transmit pulsewidth is

$$\Delta f_6 = \frac{1.21}{T} \tag{8.32}$$

The 6-dB time delay resolution width obtained by stretch processing is obtained by dividing Δf_6 by $|\alpha|$ to convert to units of time delay:

$$\tau_6 = \frac{\Delta f_6}{(B/T)} = \frac{1.21}{B} \tag{8.33}$$

Consequently, the 6-dB resolution width achieved by stretch processing is the same as that achieved by the matched filter for the LFM waveform. The 6-dB range resolution width is

$$\Delta R_6 = 1.21\frac{c}{2B} \tag{8.34}$$

Time-domain weighting is utilized in the spectral analysis processing to reduce the time sidelobes of the compressed pulse and improve the resolution performance when multiple targets are present within the range window. As an example, the use of Hamming time-domain weighting reduces the peak time sidelobe level from –13.2 dB to –42.8 dB with an increase in the 6-dB frequency resolution width to $\Delta f_6 = 1.81/T$. The 6-dB range resolution width for Hamming weighting is

$$\Delta R_6 = 1.81\frac{c}{2B} \text{ (Hamming Weighting)} \tag{8.35}$$

Range Window Width. The width of the range window is established by the bandwidth of the spectral analysis and the LFM slope of the transmit waveform. Assume a time window of width Δt and a stretch processing bandwidth B_p. A target at the edge of the time window yields a frequency offset equal to one-half of the processing bandwidth:

$$\frac{B}{T}\frac{\Delta t}{2} = \frac{B_p}{2}$$

or

$$\Delta t = T\frac{B_p}{B} = \frac{B_p}{(B/T)} \tag{8.36}$$

The range window width is

$$\Delta r = \frac{cT}{2}\frac{B_p}{B} = \frac{c}{2}\frac{B_p}{(B/T)} \tag{8.37}$$

Stretch Pulse Compression Radar Examples. This section describes three examples of radars that employ stretch pulse compression systems.

Long Range Imaging Radar.[62,63] The Long Range Imaging Radar (LRIR) is an X-band radar with stretch processing bandwidths of 0.8, 1.6 MHz, and 3.2 MHz. The wideband waveform has a swept bandwidth of 1000 MHz, to a pulsewidth of approximately 250 μs, and a LFM slope $B/T \approx 1000$ MHz/(250 μs) = 4 MHz/μs. The range window width for the 3.2 MHz processing bandwidth is

$$\Delta r = \frac{c}{2}\frac{B_p}{(B/T)} = \frac{150\,m/\mu s \times 3.2\text{ MHz}}{4\,\text{MHz}/\mu\text{s}} = 120\text{ m}$$

Millimeter Wave Radar. The stretch processing implementation for the Millimeter Wave radar (MMW) located at Kwajalein Atoll is described by Abouzahara and Avent.[64] The MMW radar operates at a carrier frequency of 35 GHz using waveforms with a maximum swept bandwidth of 1000 MHz and pulsewidth of 50 μs. The LFM slope for the transmit waveform is

$$\alpha = \frac{B}{T} = \frac{1000\text{ MHz}}{50\ \mu\text{s}} = 20\text{ MHz}/\mu\text{s}$$

The stretch processing bandwidth is $B_p = 5$ MHz. The width of the stretch processing time window is

$$\Delta t = \frac{5\text{ MHz}}{20\,\text{MHz}/\mu\text{s}} = 0.25\ \mu\text{s}$$

The reference waveform pulsewidth is $T_R = 50 + 0.25 = 50.25$ μs to avoid a loss in SNR for targets at the edges of the range window. The swept bandwidth of the reference waveform and the range window width are

$$B_R = 20\text{ MHz}/\mu\text{s} \times 50.25\ \mu\text{s} = 1005\text{ MHz}$$

$$\Delta r = \frac{c}{2}\Delta t = 150\ \text{m}/\mu\text{s} \times 0.25\ \mu\text{s} = 37.5\text{ m}$$

The 6-dB range resolution width with Hamming weighting applied over the 50-μs pulsewidth in the spectral analysis processing is

$$\Delta R_6 = 1.81\frac{c}{2B} = 1.81\frac{150\ \text{m}/\mu\text{s}}{1000\text{ MHz}} = 0.27\text{ m}$$

Cobra Dane Wideband Pulse Compression System.[67] The characteristics of the wideband pulse compression system developed for the Cobra Dane radar are summarized in Table 8.9.

TABLE 8.9 Cobra Dane Wideband Pulse Compression System Characteristics (*adapted from Filer and Hartt*[67] *© IEEE 1976*)

Characteristic	Value
Transmit LFM bandwidth	1175 to 1375 MHz
Reference LFM bandwidth	1665 to 1865 MHz*
Transmit waveform swept bandwidth, B	200 MHz
Reference waveform swept bandwidth, B_{ref}	200 MHz*
Transmit pulsewidth, T	1000 μs
Reference pulsewidth, T_{ref}	1000 μs*
Transmit waveform LFM slope	0.2 MHz/μs (up-chirp)
Compressed pulsewidth (–3 dB), τ_3	3.75 ft
Time-bandwidth product, TB	200,000
Time sidelobe level	–30 dB
Target range window	240 ft
Number of range samples	400
Range sample spacing	0.6 ft
First IF (at output of correlation mixer)	490 MHz
Second IF	60 MHz
Stretch processing bandwidth, B_p	100 kHz
A/D converter sampling frequency	1 MHz (in I and Q baseband channels)

* Excludes pulsewidth and swept bandwidth extension due to 240-ft range window

APPENDIX

Signal Analysis Summary.[68–70] Table 8.10 is a summary of signal analysis definitions and relationships. Table 8.11 shows Woodward's Fourier transform rules and pairs.[69] These relationships simplify the application of signal analysis techniques. In most cases, it will not be necessary to explicitly perform an integration to evaluate the Fourier transform or inverse Fourier transform.

TABLE 8.10 Signal Analysis Definitions and Relationships

1	Fourier transform (spectrum) of signal $x(t)$	$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi ft} dt$
2	Inverse Fourier transform of spectrum $X(f)$	$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$
3	Convolution of signals $x(t)$ and $y(t)$	$y(t) = x(t) * h(t)$ $= \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau = \int_{-\infty}^{\infty} x(t-\tau) h(\tau) d\tau$
4	Filter frequency response	$H(f) = Y(f)/X(f)$
5	Euler's identity	$e^{j\theta} = \cos\theta + j\sin\theta$

TABLE 8.10 Signal Analysis Definitions and Relationships (*Continued*)

6	Cosine and sine functions expressed in terms of complex exponentials	$\cos\theta = (e^{j\theta} + e^{-j\theta})/2$ $\sin\theta = (e^{j\theta} - e^{-j\theta})/j2$
7	Parseval's theorem (superscript asterisk indicates complex conjugate)	$\int_{-\infty}^{\infty} x(t)y^*(t)dt = \int_{-\infty}^{\infty} X(f)Y^*(f)df$ $\int_{-\infty}^{\infty} \lvert x(t)\rvert^2\, dt = \int_{-\infty}^{\infty} \lvert X(f)\rvert^2\, df$
8	rect function	$\text{rect}(t) = \begin{cases} 1, \lvert t\rvert < 1/2 \\ 0, \lvert t\rvert > 1/2 \end{cases}$
9	sinc function	$\text{sinc}(f) = \sin(\pi f)/(\pi f)$
10	Repetition operator	$\text{rep}_T[x(t)] = \sum_{n=-\infty}^{\infty} x(t - nT)$
11	Comb operator	$\text{comb}_F[X(f)] = \sum_{n=-\infty}^{\infty} X(nF)\delta(f - nF)$
12	Sampling property of delta function	$\int_{-\infty}^{\infty} x(t)\,\delta(t - t_0)dt = x(t_0)$
13	Cauchy-Schwarz inequality	$\left\lvert \int_{-\infty}^{\infty} f(x)g(x)dx \right\rvert^2 \le \int_{-\infty}^{\infty} \lvert f(x)\rvert^2\, dx \int_{-\infty}^{\infty} \lvert g(x)\rvert^2\, dx$ with equality if and only if $f(x) = k_1 g^*(x)$

Radar Transmit Waveforms.[2,68,71–73] The transmitted waveforms used in radar are bandpass signals that can be expressed in the form

$$x(t) = a(t)\cos[2\pi f_0 t + \phi(t)] \tag{8.38}$$

where $a(t)$ is the amplitude modulation (V), $\phi(t)$ is the phase modulation (rad), and f_0 is the carrier frequency (Hz). The amplitude and phase modulation functions vary slowly compared to the period of the carrier ($1/f_0$). Consequently, $x(t)$ is a narrowband waveform with a bandwidth that is small compared to the carrier frequency.

Complex Envelope. The complex envelope of $x(t)$ is given by

$$u(t) = a(t)e^{j\phi(t)} \tag{8.39a}$$

TABLE 8.11 Fourier Transform Rules and Pairs

	Signal	Spectrum	Comments
1	$x(t)$	$X(f)$	Fourier transform pair
2	$Ax(t) + Bu(t)$	$AX(f) + BU(f)$	Linearity
3	$x(-t)$	$X(-f)$	Signal time reversal
4	$x^*(t)$	$X^*(-f)$	Conjugate of signal
5	dx/dt	$j2\pi fX(f)$	Time domain differentiation
6	$-j2\pi tx(t)$	dX/df	Frequency domain differentiation
7	$x(t-\tau)$	$X(f)\exp(-j2\pi f\tau)$	Signal time shift
8	$x(t)\exp(j2\pi f_0 t)$	$X(f-f_0)$	Signal frequency shift
9	$x(t/T)$	$\|T\|X(fT)$	Time scaling
10	$x(t) * y(t)$	$X(f)Y(f)$	Time domain convolution
11	$x(t)y(t)$	$X(f) * Y(f)$	Time domain multiplication
12	$\text{rep}_T[x(t)]$	$\|1/T\|\,\text{comb}_{1/T}[X(f)]$	Woodward's repetition operator
13	$\text{comb}_T[x(t)]$	$\|1/T\|\,\text{rep}_{1/T}[X(f)]$	Woodward's comb operator
14	$X(t)$	$x(-f)$	Time-frequency interchange (duality)
15	$\delta(t)$	1	Delta function in time
16	1	$\delta(f)$	Delta function in frequency
17	$\text{rect}(t)$	$\text{sinc}(f)$	rect function in time
18	$\text{sinc}(t)$	$\text{rect}(f)$	rect function in frequency
19	$\exp(-\pi t^2)$	$\exp(-\pi f^2)$	Gaussian time function

The bandpass signal is expressed in terms of the complex envelope by

$$u(t) = \text{Re}[x(t)e^{j2\pi f_0 t}] \tag{8.39b}$$

Complex Envelope Representation of Radar Echoes.[73] The radar echo signal from a point target is

$$s_r(t) = A_r a(t - t_d)\cos[2\pi(f_0 + f_d)(t - t_d) + \phi(t - t_d)] \tag{8.40}$$

where A_r is a dimensionless amplitude scale factor, t_d is the target time delay (s), f_d is the target doppler shift (Hz), $a(t)$ is the amplitude modulation (V), $\phi(t)$ is the phase modulation (rad), and f_0 is the transmit carrier frequency (Hz). The complex envelope of $s_r(t)$ is

$$u_r(t) = A_r e^{-j2\pi f_0 t_d} u(t - t_d) e^{j2\pi f_d (t - t_d)} \tag{8.41}$$

The term $u(t - t_d)$ is the complex envelope of the transmit waveform delayed in time by t_d. The complex exponential $\exp[j2\pi f_d(t - t_d)]$ represents a linear phase modulation versus time that is impressed on the received echo signal by the doppler shift f_d. The carrier phase shift is $\theta_c = -2\pi f_0 t_d$.

The time delay and doppler shift are expressed in terms of target range and range-rate by $t_d = 2R/c$ (s) and $f_d = -(2/\lambda)V_r$ (Hz), where R is the target range (m), $V_r = dR/dt$

is the range-rate (negative for an incoming target), c is the speed of light, and $\lambda = c/f_0$ (m) is the carrier wavelength.

Matched Filters.[2,74] A matched filter achieves maximum output signal-to-noise ratio for a signal received in white noise. The matched filter frequency response for a signal $u(t)$ is

$$H_{\text{mf}}(f) = k_1 U^*(f) e^{-j2\pi f t_1} \tag{8.42}$$

where k_1 is an arbitrary complex constant and $U(f)$ is the spectrum of $u(t)$. The time delay t_1 is required to exceed the duration of $u(t)$ to achieve a causal impulse response that is zero for negative time. The matched filter impulse response is

$$h_{\text{mf}}(t) = k_1 u^*(t_1 - t) \tag{8.43}$$

The peak signal-to-noise to mean-noise-power ratio at the output of a filter with frequency response $H(f)$ is defined as

$$(S/N)_o = \frac{A_0^2}{\sigma_{n_o}^2} \tag{8.44}$$

where A_o is the matched filter output signal amplitude at the peak of the signal and σ^2_{no} is the matched filter output noise power. The matched filter output SNR is given by*

$$(S/N)_{\text{mf}} = \frac{2E}{N_0} \tag{8.45}$$

where E is the energy of the received bandpass signal at the matched filter input (J), and N_0 is the one-sided noise power spectrum at the matched filter input (W/Hz).

Filter Matching Loss. Filter matching loss L_m is the loss in SNR that results when a signal is not processed using a matched filter. The filter matching loss is defined as

$$L_m = \frac{(S/N)_{\text{mf}}}{(S/N)_o} \tag{8.46}$$

where $(S/N)_{\text{o}}$ is the SNR at the output of a filter with frequency response $H(f)$ and $(S/N)_{\text{mf}}$ is the matched filter SNR. The filter matching loss can also be expressed as

$$L_m = \frac{(2E/N_0)}{(S/N)_o} \tag{8.47}$$

where the matched filter SNR is given by $(S/N)_{\text{mf}} = (2E/N_0)$. The filter matching loss is ≥ 1, where $L_m = 1$ for the matched filter. The filter matching loss expressed in decibels is $L_m(\text{dB}) = 10\log(L_m)$ and equals 0 dB for the matched filter.

---

* An alternate definition of signal-to-noise ratio is also used in the literature in which the signal power at the peak of the waveform is averaged over one cycle of the carrier.[75,76] In this case, the average signal power is one-half of the peak signal power and the matched-filter output SNR is E/N_0.

Ambiguity Functions.[2,71,72,77–79] The autocorrelation† function for a transmit waveform with complex envelope $u(t)$ is defined as

$$\chi_u(\tau, f_d) = \int_{-\infty}^{\infty} u(t)u^*(t+\tau)e^{j2\pi f_d t}dt \tag{8.48}$$

where τ is the relative time delay and f_d is doppler shift. The relative time delay is positive for a target further in range than a reference target, and doppler frequency is positive for an incoming target (negative range-rate). The complex envelope $u(t)$ is normalized to unit energy

$$\int_{-\infty}^{\infty} |u(t)|^2 dt = 1 \tag{8.49}$$

The ambiguity function of $u(t)$ is defined as the square magnitude of the autocorrelation function

$$\Psi_u(\tau, f_d) = |\chi_u(\tau, f_d)|^2 \tag{8.50}$$

The ambiguity function is interpreted as a surface above the delay-doppler ($\tau - f_d$) plane. The maximum value of the ambiguity function is unity at the origin ($\tau = f_d = 0$):

$$\Psi_u(\tau, f_d) \le \Psi_u(0,0) = 1 \tag{8.51}$$

The volume under the ambiguity surface is unity for any waveform $u(t)$:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Psi_u(\tau, f_d)\, d\tau df_d = 1 \tag{8.52}$$

In the general case, where the energy of the complex envelope is not normalized to unity, the value of the ambiguity function at the origin is equal to $(2E)^2$ where E is the energy of the bandpass signal corresponding to $u(t)$, and the volume under the ambiguity function is also equal to $(2E)^2$. The normalization condition is equivalent to the assumption that the energy of the bandpass transmit waveform equals 0.5 J.

Matched Filter Time Response. The matched filter time response to a target with doppler shift f_d can be expressed in terms of the autocorrelation function. The matched filter impulse response with $k_1 = 1$ and $t_1 = 0$ is

$$h_{\text{mf}}(t) = u^*(-t) \tag{8.53}$$

The matched filter input signal is assumed to have zero time delay and a doppler shift f_d:

$$s(t) = u(t)e^{j2\pi f_d t} \tag{8.54}$$

† The terminology for this function is not standardized in the literature. Woodward[69] uses the term correlation function. The term time-frequency autocorrelation function is used by Spafford.[80] The signs associated with τ and fd within the integrand also differ in the literature. The standardized definition proposed by Sinsky and Wang[78] is used in this chapter.

The matched filter output signal $y(t)$ is found by convolution of $s(t)$ with the matched filter impulse response $h_{\text{mf}}(t)$:

$$y(t) = \int_{-\infty}^{\infty} u(t')u^*(t'-t)e^{j2\pi f_d t'}dt' \tag{8.55}$$

Comparison of this result with the definition of the autocorrelation function shows that the matched filter response can be expressed as

$$y(t) = X_u(-t, f_d) \tag{8.56}$$

As a result, the matched filter time response for a target with doppler frequency f_d is a time-reversed version of the autocorrelation function.

Conditions for Target Resolution in Time Delay and Doppler Frequency.[72,78] Assume that two targets with equal radar cross sections are present at the same angular position. The first target (termed the *reference target*) is located at the origin of the delay-doppler plane with zero relative time delay and zero doppler frequency, and the second target is at relative time delay τ and doppler frequency f_d. The relative time delay is positive when the second target is farther in range than the reference target and the doppler frequency is positive for an incoming target. The matched-filter output power for the reference target is proportional to the ambiguity function and is given by

$$P_{\text{ref}} = \Psi_u(0,0) = 1 \tag{8.57}$$

The matched filter output power for the second target, evaluated at the peak of the reference target, is

$$P_2 = \Psi_u(\tau, f_d) \tag{8.58}$$

The second target is unresolved from the reference target at locations in the delay-doppler plane where $\Psi_u(\tau, f_d) \approx 1$.

REFERENCES

1. J. R. Klauder, A. C. Price, S. Darlington, and W. J. Albersheim, "The theory and design of chirp radars," *Bell Syst. Tech. J.*, vol. 39, pp. 745–808, July 1960.
2. C. E. Cook and M. Bernfield, *Radar signals: An Introduction to Theory and Application*, New York: Academic Press, 1967.
3. C. E. Cook and J. Paolillo, "A pulse compression predistortion function for efficient sidelobe reduction in a high-power radar," *Proc. IEEE*, pp. 377–389, April 1964.
4. T. T. Taylor, "Design of line-source antennas for narrow beamwidth and low sidelobes," *IRE Trans.*, vol. AP-3, pp. 16–28, January 1955.
5. R. C. Hansen, "Aperture theory," in *Microwave Scanning Antennas*, vol. I, R. C. Hansen (ed.), New York: Academic Press, 1964, chap. 1.
6. H. Gautier and P. Tournois, "Signal processing using surface-acoustic-wave and digital components," *IEEE Proc.*, vol. 127, pt. F, pp. 92–93, April 1980.
7. A. J. Slobodnik, Jr., "Surface acoustic waves and SAW materials," *Proc. IEEE*, vol. 64, pp. 581–594, May 1976.
8. T. W. Bristol, "Acoustic surface-wave-device applications," *Microwave J.*, vol. 17, pp. 25–27, January 1974.

9. J. W. Arthur, "Modern SAW-based pulse compression systems for radar applications," *Electronics & Communications Engineering Journal*, December 1995.
10. R. C. Williamson, "Properties and applications of reflective-array devices," *Proc. IEEE*, vol. 64, pp. 702–703, May 1976.
11. G. W. Judd, "Technique for realizing low time sidelobe levels in small compression ratio chirp waveforms," *Proc. IEEE Ultrasonics Symp*., 1973, pp. 478–481.
12. A. Pohl, C. Posch, F. Seifert, and L. Reindl, "Wideband compressive receiver with SAW convolver," *1995 IEEE Ultrasonics Symposium*, pp. 155–158.
13. X. Shou, J. Xu, H. Wang, and Q. Xu, "SAW pulse compression systems with lower sidelobes," *1997 Asia Pacific Microwave Conference*, pp. 833–835.
14. T. Murakami, "Optimum waveform study for coherent pulse doppler," *RCA Final Rept*., prepared for Office of Naval Research, Contract Nonr 4649(00)(x), AD641391, February 28, 1965.
15. T. Collins and P. Atkins, "Nonlinear frequency modulation chirps for active sonar," *IEE Proc.-Radar, Sonar Navig*., vol. 146, no. 6, pp. 312–316, December 1999.
16. L. R. Varschney and D. Thomas, "Sidelobe reduction for matched range processing," *2003 IEEE Radar Conference*, pp. 446–451.
17. N. Levanon and E. Mozeson, *Radar Signals*, New York: IEEE Press, John Wiley & Sons, Inc., 2004, pp. 106, 145.
18. R. H. Barker, "Group synchronization of binary digital systems," *in Communication Theory*, W. Jackson (ed.), New York: Academic Press, 1953, pp. 273–287.
19. P. J. Edmonson, C. K. Campbell, and S. F. Yuen, "Study of SAW pulse compression using 5×5 Barker codes with quadriphase IDT geometries," *1988 IEEE Ultrasonics Symposium*, pp. 219–222.
20. T. Felhauer, "Design and analysis of new P(n,k) polyphase pulse compression codes," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 30, no. 3, pp. 865–874, July 1994.
21. R. Turyn and J. Stover, "On binary sequences," *Proc. Am. Math. Soc*., vol. 12, pp. 394–399, June 1961.
22. D. G. Luenburger, "On Barker codes of even length," *Proc. IEEE*, vol. 51, pp. 230–231, January 1963.
23. R. Turyn, "On Barker codes of even length," *Proc. IEEE* (correspondence), vol. 51, p. 1256, September 1963.
24. L. Bomer and M. Antweiler, "Polyphase Barker sequences," *Electronics Letters*, vol. 23, no. 23, pp. 1577–1579, November 9, 1989.
25. H. Meikle, *Modern Radar Systems*, Norwood, MA: Artech House, 2001, p. 258.
26. B. L. Lewis, "Range-time-sidelobe reduction technique for FM-derived polyphase PC codes," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 29, no.3, pp. 834–863, July 1993.
27. W. W. Peterson and E. J. Weldon, Jr., *Error Correcting Codes*, Cambridge: M.I.T. Press, 1972, app. C.
28. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw Hill, 2001, p. 367.
29. N. Levanon and E. Mozeson, *Radar Signals*, New York: IEEE Press, John Wiley & Sons, Inc., 2004, pp. 106–109.
30. L. Bömer and M. Antweiler, "Polyphase Barker sequences," *Electronics Letters*, vol. 25, no. 23, pp. 1577–79, November 9, 1989.
31. W-D. Wirth, *Radar Techniques Using Array Antennas*, IEE Radar, Sonar, Navigation and Avionics Series 10, London: The Institution of Electrical Engineers, 2001.
32. R. L. Frank, "Polyphase codes with good nonperiodic correlation properties," *IEEE Trans*., vol. IT-9, pp. 43–45, January 1963.
33. B. L. Lewis and F. F. Kretschmer, Jr., "A new class of polyphase pulse compression codes and techniques," *IEEE Trans*., vol. AES-17, pp. 364–372, May 1981. (See correction, IEEE Trans., vol. AES-17, p. 726, May 1981.)
34. B. L. Lewis, "Range-time-sidelobe reduction technique for FM-derived polyphase PC codes," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 29, no.3, pp. 834–863, July 1993.

35. B. L. Lewis and F. F. Kretschmer, Jr., "Linear Frequency Modulation Derived Polyphase Pulse Compression Codes," *IEEE Trans. on Aerospace and Electronics Systems*, AES-18, no. 5, pp. 636–641, September 1982.
36. J. W. Taylor and H. J. Blinchikoff, "Quadriphase code-a radar pulse compression signal with unique characteristics," *IEEE Trans. Aerospace and Electronic Systems*, vol. 24, no. 2, pp. 156–170, March 1988.
37. H. J. Blinchikoff, "Range sidelobe reduction for the quadriphase codes," IEEE Trans. Aerospace and Electronic Systems, vol. 32, no. 2, April 1996, pp. 668–675.
38. N. Levanon, "Stepped-frequency pulse-train radar signal," *IEE Proc-Radar Sonar Navigation*, vol. 149, no. 6, December 2002.
39. A. W. Rihaczek, *Principles of High-Resolution Radar*, New York: McGraw-Hill Book Company, 1969, chap. 8.
40. J. P. Donohue and F. M. Ingels, "Ambiguity function properties of frequency hopped radar/sonar signals," *Proc. of the 1989 Southeastcon*, session 10B6, pp. 85–89.
41. J. P. Costas, "A study of a class of detection waveforms having nearly ideal range-doppler ambiguity properties," *Proc. of the IEEE*, vol. 72, no. 8, August 1984.
42. B. R. Mahafza, *Radar Systems Analysis and Design using MATLAB, 2000*, Boca Raton: Chapman & Hall/CRC, 2000.
43. B. M. Popvik, "New construction of Costas sequences," *Electronic Letters*, vol. 25, no.1, January 5, 1989.
44. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw Hill, 2001, pp. 355–57.
45. D. P. Morgan, "Surface acoustic wave devices and applications," *Ultrasonics*, vol. 11, pp. 121–131, 1973.
46. L. O. Eber and H. H. Soule, Jr., "Digital generation of wideband LFM waveforms," *IEEE Int. Radar Conf. Rec*., 1975, pp. 170–175.
47. AD9858 1-GSPS direct digital synthesizer data sheet, Rev. A, 2003, Analog Devices, Norwood, MA (available at www.analog.com).
48. J. K. Hartt and L. F. Sheats, "Application of pipeline FFT technology in radar signal and data processing," *EASCON Rec*., 1971, pp. 216–221; reprinted in David K. Barton, *Radars*, vol. 3, Ann Arbor: Books on Demand UMI, 1975.
49. P. E. Blankenship and E. M. Hofstetter, "Digital pulse compression via fast convolution," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-23, no. 2, pp. 189–222, April 1975.
50. L. W. Martinson and R. J. Smith, "Digital matched filtering with pipelined floating point fast Fourier transforms (FFTs)," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-23, no. 2, pp. 222–234, April 1975.
51. L. E. Pellon, "A double Nyquist digital product detector for quadrature sampling," *IEEE Trans. on Signal Processing*, vol. 40, no. 7, pp. 1670–1680.
52. G. A. Shaw and S. C. Pohlig, "I/Q baseband demodulation in the RASSP SAR benchmark," Project Report RASSP-4, Massachusetts Institute of Technology, Lincoln Laboratory, 25 August 1995, http://www.ll.mit.edu/llrassp/documents.html.
53. M. A. Richard, "Digital I/Q," Section 3.7.3 in *Fundamentals of Radar Signal Processing*, New York: McGraw-Hill, 2005.
54. L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975, chap. 2.
55. J. J. Gostin, "The GE592 solid-state radar," *IEEE EASCON 1980*, pp. 197–203.
56. E. L, Cole, P. A DeCesare, M. J. Martineaus, R. S. Baker, and S. M. Buswell, "ASR-12: A next generation solid-state air traffic control radar," *IEEE 1998 Radar Conference*, pp. 9–14.
57. W. J. Caputi, Jr., "Stretch: A time-transformation technique," *IEEE Trans*., vol. AES-7, pp. 269–278, March 1971.
58. W. J. Caputi, "A technique for the time-transformation of signals and its application to directional systems," *The Radio and Electronic Engineer*, pp. 135–142, March 1965.

59. W. J. Caputi, "Swept-heterodyne apparatus for changing the time-bandwidth product of a signal," U.S. Patent 3283080, November 1, 1966.
60. W. J. Caputi, "Pulse-type object detection apparatus," U.S. Patent, November 21, 1967.
61. K. R. Roth, M. E. Austin, D. J. Frediani, G. H. Knittel, and A. V. Mrstik, "The Kiernan reentry measurements system on Kwajalein Atoll," The Lincoln Laboratory Technical Journal, vol. 2, no. 2, 1989.
62. D. R. Bromaghim and J. P. Perry, "A wideband linear fm ramp generator for the long-range imaging radar," *IEEE Trans, Microwave Theory and Techniques*, vol. MTT-26, no. 5, pp. 322–325, May 1978.
63. G. R. Armstrong and M. Axelbank, "Description of the long-range imaging radar," Project Report PSI-85, Massachusetts Institute of Technology, Lincoln Laboratory, November 16, 1977.
64. M. D. Abouzahra and R. K. Avent, "The 100-kW millimeter-wave radar at the Kwajalein Atoll," *IEEE Antennas and Propagation Magazine*, vol. 36, no. 2, pp. 7–19, April 1994.
65. W. C. Kellog, "Digital processing rescues hardware phase errors," *Microwaves & RF*, pp. 63–67, 80, November 1982.
66. C. L. Temes, "Sidelobe suppression in a range channel pulse-compression radar," *IRE Trans*., vol. MIL-6, pp. 162–169, April 1962.
67. E. Filer and J. Hartt, "COBRA DANE wideband pulse compression system," *IEEE EASCON* '76, 1976, pp. 61-A–61-M.
68. S. Stein and J. J. Jones, "*Modern Communication Principles with Application to Digital Signaling*," New York: McGraw-Hill, 1967.
69. P. M. Woodward, *Probability and Information Theory with Application to Radar*, Pergamon Press, 1960.
70. D. Brandwood, "Fourier Transforms," in *Radar and Signal Processing*, Boston: Artech House, 2003.
71. G. W. Deley, "Waveform design," in *Radar Handbook*, M. I. Skolnik (ed.), 1st Ed., New York: McGraw-Hill, 1970.
72. A. I. Sinsky, "Waveform selection and processing" in *Radar Technology*, E. Brookner (ed.), Boston: Artech House, 1977, Chap. 7.
73. C. W. Helstrom, *Statistical Theory of Signal Detection*, 2nd Ed., Pergamon Press, 1968.
74. G. L. Turin, "An introduction to matched filters," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 311–329, June 1960.
75. D. K. Barton, *Modern Radar System Analysis and Modeling*, Canton, MA: Artech House Inc, 2005, Chap. 5, p. 197.
76. F. E. Nathanson, J. P. Reilly, and M. N. Cohen, *Radar Design Principles: Signal Processing and the Environment*, 2nd Ed. New York: McGraw-Hill, 1991. chap. 8, p. 357.
77. A. W. Rihaczek, "Radar signal design for target resolution," *Proc. IEEE*, vol. 53, pp. 116–128, February 1965.
78. A. I. Sinsky and C. P. Wang, "Standardization of the definition of the ambiguity function," *IEEE Trans. Aerospace and Electronic Systems*, pp. 532–533, July 1974.
79. IEEE standard radar definitions, IEEE Std 686-1990, The Institution of Electrical and Electronic Engineers, New York, NY, 1990. The ambiguity function is defined on page 55 using the standardized definition given by Sinsky and Wang.
80. L. J. Spafford, "Optimum radar signal processing in clutter," *IEEE Trans. Information Theory*, vol. IT-14, no. 5, pp. 734–743, September 1968.

Chapter 9

Tracking Radar

Dean D. Howard

Consultant to ITT Industries, Inc.

9.1 INTRODUCTION

Typical tracking radars have a pencil beam to receive echoes from a single target and track the target in angle, range, and/or doppler. Its resolution cell—defined by its antenna beamwidth, transmitter pulse length (effective pulse length may be shorter with pulse compression), and/or doppler bandwidth—is usually small compared with that of a search radar and is used to exclude undesired echoes or signals from other targets, clutter, and countermeasures. Electronic beam-scanning phased array radars may track multiple targets by sequentially dwelling upon and measuring each target while excluding other echo or signal sources.

Because of its narrow beamwidth, typically from a fraction of 1° to 1 or 2°, tracking radars usually depend upon information from a surveillance radar or other source of target location to acquire the target, i.e., to place its beam on or in the vicinity of the target before initiating a track. Scanning of the beam within a limited angle sector may be needed to acquire the target within its beam and center the range-tracking gates on the echo pulse prior to locking on the target or closing the tracking loops. The gate acts like a fast-acting on-off switch that turns the receiver "on" at the leading edge of the target echo pulse and "off" at the end of the target echo pulse to eliminate undesired echoes. The range-tracking system performs the task of keeping the gate centered on the target echo, as described in Section 9.5.

The primary output of tracking radar is the target location determined from the pointing angles of the beam and position of its range-tracking gates. The angle location is the data obtained from synchros and encoders on the antenna tracking axes (or data from a beam-positioning computer on an electronic-scan phased array radar). In some cases, tracking lag is measured by converting tracking-lag-error voltages from the tracking loops to units of angle. This data is used to add to or subtract from the angle shaft position data for real-time correction of tracking lag.

There are a large variety of tracking-radar systems, including some that achieve simultaneously both surveillance and tracking functions. A widely used type of tracking radar and the one discussed in detail in this chapter is a ground-based system consisting of a pencil-beam antenna mounted on a rotatable platform with servo motor drive of its azimuth and elevation position to follow a target (Figure 9.1*a*). Errors in pointing direction are determined by sensing the angle of arrival of the echo wavefront and corrected by positioning the antenna to keep the target centered

in the beam. Modern requirements for simultaneous precision tracking of multiple targets has driven the development of the electronic scan array monopulse radar with the capability to switch its beam pulse-to-pulse among multiple targets. The AN/MPS-39 shown in Figure 9.1*b* is an example of a highly versatile electronic scan monopulse missile-range instrumentation radar.

The principal applications of precision tracking radar are weapon control and missile-range instrumentation. In both applications, a high degree of precision and an accurate prediction of future position of the target are generally required. The earliest use of tracking radar was gunfire control. The azimuth angle, elevation angle, and the range to the target were measured, and from the rate of change of these parameters, the velocity vector of the target (speed and direction) was computed and its future position predicted. This information was used to move the gun to lead the target and set the fuze delay. The tracking radar performs a similar role in providing guidance information and steering commands for missile control.

In missile-range instrumentation, the tracking-radar output is used to measure the trajectory of the missile and to predict future position. Tracking radars are used to compute the impact point of a launched missile continuously during the launch phase in case of missile failure for range safety. If the impact point approaches a populated or other critical area, the missile is destroyed. Missile-range instrumentation radars are normally used with a beacon (pulse repeater) to provide a point-source echo—usually its pulse is delayed to separate it from the target echo—and with high signal-to-noise ratio, to achieve precision tracking on the order of 0.05 mil in angle and 5 m in range.

(a) (b)

FIGURE 9.1 (*a*) AN/FPQ-6 C-band monopulse precision tracking radar installation at the NASA Wallops Island Station, VA. It has a 29-ft-diameter dish and a specified tracking precision of 0.05 mrad rms. (*b*) AN/MPS-39 C-band electronic scan phased-array Multi Object Tracking Radar (MOTR) installed at the White Sands Missile Range. (*Photo of the AN/MPS-39 courtesy of the White Sands Missile Range and Lockheed Martin.*)

This chapter describes the monopulse (simultaneous lobing with either phase comparison or amplitude comparison), conical-scan, and sequential lobing tracking-radar techniques with the main emphasis on the amplitude-comparison monopulse (simultaneous lobing) radar.

9.2 MONOPULSE (SIMULTANEOUS LOBING)

The susceptibility of conical scanning and sequential lobing tracking techniques to echo amplitude fluctuations and amplitude jamming (as described in Section 9.3) was the major reason for the development of tracking radar that provides simultaneously all the necessary lobes for angle-error sensing. This required that the output from the lobes be compared simultaneously on a single pulse, eliminating the effects of echo amplitude change with time. The technique to accomplish this was initially called *simultaneous lobing,* which was descriptive of the technique. Later, the term *monopulse* was coined, referring to the ability to obtain angle error information on a single pulse. It has become the commonly used name for this tracking technique; even though, the lobes are generated simultaneously and *monopulse* tracking can be performed with CW radar.

The original monopulse tracking radars suffered in antenna efficiency and complexity of microwave circuitry because waveguide signal-combining circuitry was a relatively new art. These problems were overcome and monopulse radar, with modern compact off-the-shelf processing circuitry, can readily outperform scanning and lobing systems. The monopulse technique also has an inherent capability for high-precision angle measurement because its feed structure is compact with short signal paths and rigidly mounted with no moving parts. This has made possible the development of pencil-beam tracking radars that meet missile-range instrumentation-radar requirements of 0.003° angle-tracking precision.

This chapter is devoted to tracking radar, but monopulse techniques are used in other systems including homing devices, direction finders, and some search radars. However, most of the basic principles and limitations of monopulse apply for all applications. More general coverage is found in Sherman[1] and Leonov and Formichev.[2]

Amplitude-Comparison Monopulse. A method for visualizing the operation of an amplitude-comparison receiver is to consider the echo signal at the focal plane of an antenna.[3] The echo is focused to a finite size "spot." The "spot" is centered on the focal plane when the target is on the antenna axis and moves off center when the target moves off axis. The antenna feed is located at the focal point to receive maximum energy from a target on axis.

The amplitude-comparison feed is designed to sense any feed plane displacement of the spot from the center of the focal plane. A monopulse feed using the four-horn square, for example, would be centered at the focal plane. It provides symmetry so that when the spot is centered, equal energy falls on each of the four horns. The radar senses target displacement from the antenna axis that shifts the spot off of the center of the focal plane by measuring the resultant unbalance of energy received in the four horns. This is accomplished by use of microwave waveguide hybrids to subtract outputs of pairs of horns, providing a sensitive device that gives signal output when there is an unbalance caused by the target being off axis. The RF circuitry for a conventional four-horn square feed (see Figure 9.2) subtracts the output of

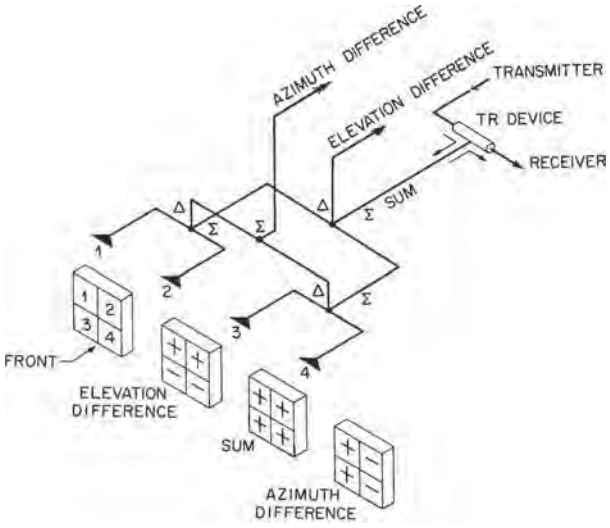


FIGURE 9.2 Microwave-comparator circuitry used with a four-horn monopulse feed

the left pair from the output of the right pair to sense any unbalance in the azimuth direction. It also subtracts the output of the top pair from the output of the bottom pair to sense any unbalance in the elevation direction. In addition, the circuitry adds the output of all four horns for a sum signal for detection, monopulse processing, and range tracking.

The comparator shown in Figure 9.2 is the circuitry that performs the addition and subtraction of the feed horn outputs to obtain monopulse sum and difference signals. It is illustrated with hybrid-T (or magic-T) waveguide components. These are four-port devices that, in basic form, have the inputs and outputs located at right angles to each other. However, the magic T's have been developed in convenient "folded" configurations for a very compact comparator. The performance of these and other similar four-port devices is described in Chapter 4 of Sherman.[1]

The subtractor outputs are called *difference signals,* which are zero when the target is on axis, increasing in amplitude with increasing displacement of the target from the antenna axis. The difference signals also change 180° in phase from one side of center to the other. The sum of all four-horn outputs provides a reference signal to control angle-tracking sensitivity (volts per degree of error) to remain constant, even though the target echo signal may vary over a large dynamic range. This is accomplished by automatic gain control (AGC) to keep the sum signal output and angle-tracking loop gains constant for stable automatic angle tracking.

Figure 9.3 is a block diagram of typical monopulse radars. The sum signal, elevation difference signal, and azimuth difference signal are each converted to intermediate frequency (IF), using a common local oscillator to maintain relative phase at IF. The IF sum-signal output is detected and provides the video input to the range tracker. The range tracker measures and tracks the time of arrival of the desired target echo and provides gate pulses that turn on the radar receiver channels only during the brief period when the desired echo is expected. The gated video is used to generate the dc

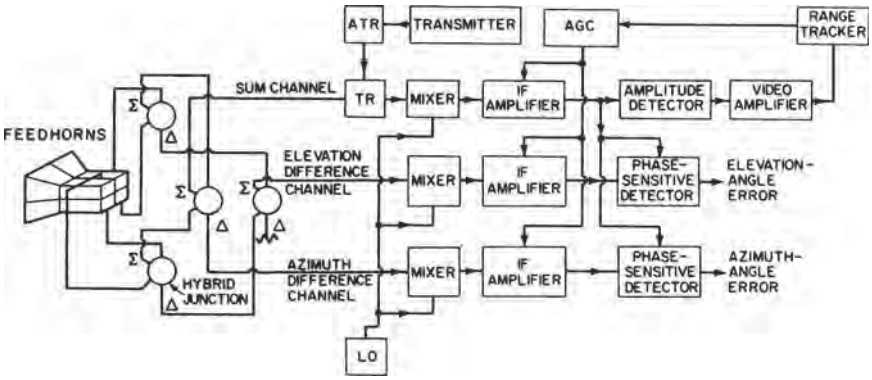


FIGURE 9.3 Block diagram of a conventional monopulse tracking radar

voltage proportional to the magnitude of the Σ signal or $|\Sigma|$ for the AGC of all three IF amplifier channels. The AGC maintains constant angle-tracking sensitivity (volts per degree error), even though the target echo signal varies over a large dynamic range, by controlling gain or dividing by $|\Sigma|$. AGC is necessary to keep the gain of the angle-tracking loops constant for stable automatic angle tracking. Some monopulse systems, such as the two-channel monopulse, can provide instantaneous AGC or normalizing by use of log detectors as described later in this section.

The sum signal at the IF output also provides a reference signal to phase detectors that derive angle-tracking-error voltages from the difference signals. The phase detectors are essentially dot-product devices producing the output voltage

$$e = \frac{|\Sigma|\ |\Delta|}{|\Sigma|\ |\Delta|} \cos\theta \qquad \text{or} \qquad e = \frac{\Delta}{|\Sigma|} \cos\theta \tag{9.1}$$

where e = angle-error-detector output voltage
$|\Sigma|$ = magnitude of sum signal
$|\Delta|$ = magnitude of difference signal
θ = phase angle between sum and difference signals

The dot-product error detector is only one of a wide variety of monopulse angle-error detectors described in Chapter 7 of Sherman.[1]

Normally, θ is either 0° or 180° when the radar is properly adjusted, and the only purpose of the phase-sensitive characteristic is to provide a plus or minus polarity corresponding to $\theta = 0°$ and $\theta = 180°$, respectively, giving a + or – polarity to the angle-error-detector output to indicate to the servo which direction to drive the pedestal.

In a pulsed tracking radar, the angle-error-detector output is bipolar video; that is, it is a video pulse with an amplitude proportional to the angle error and whose polarity (positive or negative) corresponds to the direction of the error. This video is typically processed by a sample and hold circuit that charges a capacitor to the peak video-pulse voltage and holds the charge until the next pulse, at which time the capacitor is discharged and recharged to the new pulse level. With moderate low-pass filtering, this gives the dc error voltage output to the servo amplifier to correct the antenna position.

The three-channel amplitude-comparison monopulse tracking radar is the most commonly used monopulse system. However, the three signals may sometimes be combined in other ways to perform with a two-channel receiver system (as described later in this section) used in some current surface-to-air missile (SAM) systems.

Monopulse-Antenna Feed Techniques. Monopulse-radar feeds may have any of a variety of configurations. Single apertures are also employed by use of higher-order waveguide modes to extract angle-error-sensing difference signals. There are many tradeoffs in feed design because optimum sum and difference signals, low sidelobe levels, selectable polarization capability, and simplicity cannot all be fully satisfied simultaneously. The term *simplicity* refers not only to cost savings but also to the use of noncomplex circuitry, which is necessary to provide a broadband system with good boresight stability to meet precision-tracking requirements. (*Boresight* is the electrical axis of the antenna or the angular location of a signal source within the antenna beam at which the angle-error-detector outputs go to zero.)

Some of the typical monopulse feeds are described to show the basic relations and tradeoffs involved in the various performance factors and how the more important factors can be optimized by a feed configuration but at the price of lower performance in other areas.[4] Many new techniques have been added since the original four-horn square feed in order to provide good or excellent performance in all desired feed characteristics in a well-designed monopulse radar.

The original four-horn square monopulse feed is inefficient because the optimum feed size aperture for the difference signals is approximately twice the optimum size for the sum signal.[5] Consequently, an intermediate size is typically used with a significant compromise for both sum and difference signals. The optimum four-horn square feed, which is subject to this compromise, described in Sherman,[1] is based on minimizing the angle error caused by receiver thermal noise. However, if sidelobes are a prime consideration, a somewhat different feed size may be desired.

The limitation of the four-horn squared feed is that the sum- and difference-signal *E* fields cannot be controlled independently. If independent control could be provided, the *ideal* would be approximately as described in Figure 9.4, with twice the dimension for the difference signals in the plane of error sensing than for the sum signal.[5]

A technique used by the MIT Lincoln Laboratory to approach the ideal is a 12-horn feed (Figure 9.5). The overall feed, as illustrated, is divided into small parts, and the microwave circuitry selects the portions necessary for the sum and difference signals to approach the ideal. One disadvantage is that this feed requires a very complex microwave circuit. Also, the divided four-horn portions of the feed are each four element arrays that generate large feed sidelobes in the *H*-plane because of the double-peak *E* field. Another consideration is that the 12-horn feed is not practical for focal-point-fed parabolas or reflectarrays because of its size. A focal-point feed is usually small to produce a broad pattern and must be compact to avoid blockage of the antenna aperture. In some cases, the small optimum size required is below waveguide cutoff, and dielectric loading of the horn apertures becomes necessary to avoid cutoff.

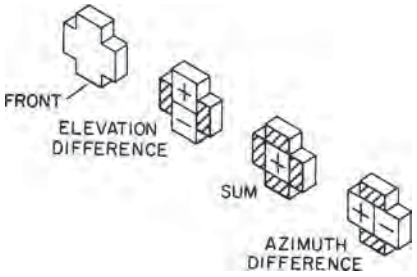


FIGURE 9.4 Approximately ideal feed-aperture *E*-field distribution for sum and difference signals

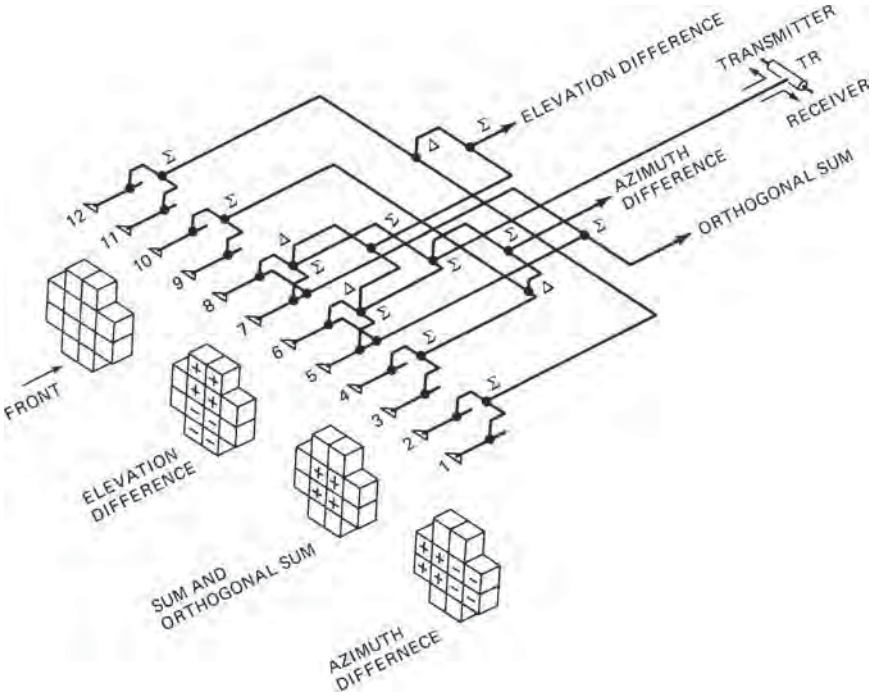


FIGURE 9.5 Twelve-horn feed

A practical approach to monopulse feed design uses higher-order waveguide modes rather than multiple horns for independent control of sum- and difference-signal *E* fields. This allows much greater simplicity and flexibility. A triple-mode two-horn feed used by RCA[5,6] retracts the *E*-plane septa to allow both the TE_{10} and TE_{30} modes to be excited and propagate in the double-width septumless region, as illustrated in Figure 9.6. At the septum, the double-humped *E* field is represented by the combined TE_{10} and TE_{30} modes subtracting at the center and adding at the TE_{30}-mode outer peaks. However, because the two modes propagate at different velocities, a point is reached farther down the double-width guide where the two modes add in the center and subtract at the outer humps of the TE_{30} mode. The result is a sum-signal *E* field concentrated, as desired, toward the center of the feed aperture.

This shaping of the sum-signal *E* field is accomplished independently of the difference-signal *E* field. The difference signal is two TE_{10}-mode signals, side by side, arriving at the septum of Figure 9.6 out of phase. At the septum, it becomes the TE_{20} mode, which propagates to the horn aperture and uses the full width of the horn as desired. The TE_{20} mode has zero *E* field in the center of the waveguide where the septum is located and is unaffected by the septum.

A further step in feed development is the four-horn triple-mode feed illustrated in Figure 9.7.[5] This feed uses the same approach as described above but with the addition of a top and bottom horn. This allows the *E*-plane difference signal to couple to all four horns and uses the full height of the feed. The sum signal uses only the center two horns to limit its *E* field in the *E*-plane as desired for the ideal field shaping.

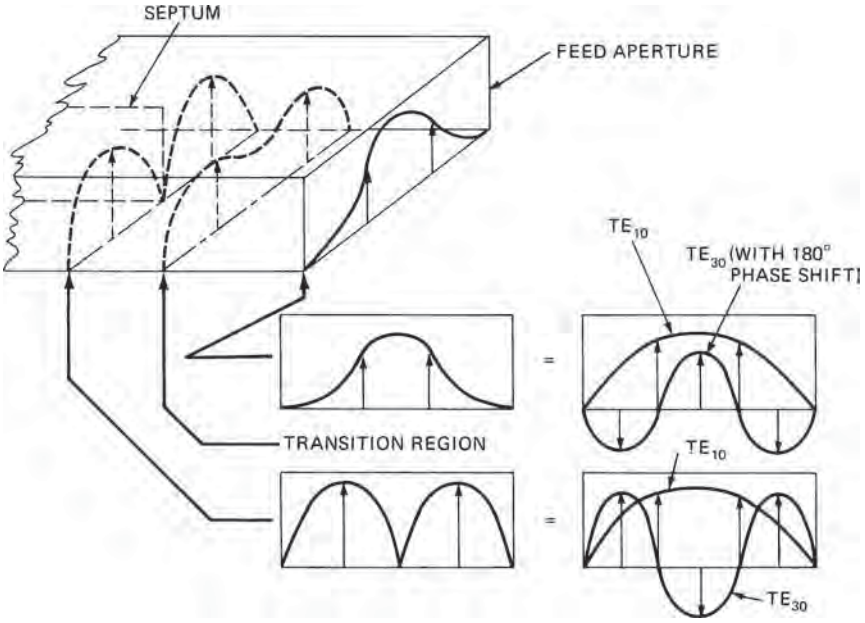


FIGURE 9.6 Use of retracted septum to shape the sum-signal *E* field

The use of smaller top and bottom horns is a simpler method of concentrating the *E*-field toward the center of the feed, where the full horn width is not needed.

The feeds described thus far are for linear-polarization operation. When circular polarization is needed in a paraboloid-type antenna, square or circular cross-section horn throats are used. The vertical and horizontal components from each horn are

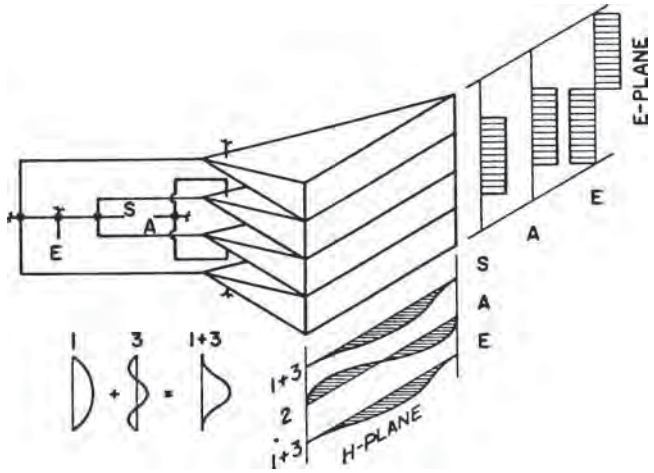


FIGURE 9.7 Four-horn triple-mode feed (*after P. W. Hannan*[5] © *IEEE 1961*)

separated and comparators provided for each polarization. The sum and difference signals from the comparators are combined with 90° relative phase to obtain circular polarization. Use of the previously described feeds for circular polarization would require the waveguide circuitry to be prohibitively complex. Consequently, a five-horn feed has been used as illustrated in Figure 9.8.

The five-horn feed is selected because of the simplicity of the comparator that requires only two magic (or hybrid) T's for each polarization. The sum and difference signals are provided for the two linear-polarization components and, in an AN/FPQ-6 radar, are combined in a waveguide switch for selecting polarization. The switch selects either the vertical or the horizontal input component or combines them with a 90° relative phase for circular polarization. This feed does not provide optimum sum- and difference-signal *E* fields because the sum horn occupies space desired for the difference signals. Generally, an undersized sum-signal horn is used as a compromise. However, the five-horn feed is a practical choice between complexity and efficiency. It has been used in several instrumentation radars including the AN/FPQ-6, AN/FPQ-10, AN/TPQ-18, andAN/MPS-36[7,8] and in the AN/TPQ-27 tactical precision-tracking radar.[9]

The multimode feed technique can be expanded to other higher-order modes for error sensing and *E*-field shaping.[10,11,12] The difference signals are contained in unsymmetrical modes such as the TE_{20} mode for *H*-plane error sensing and combined TE_{11} and TM_{11} modes for *E*-plane error sensing. These modes provide the difference signals, and no comparators are used.[10] Generally, mode coupling devices can give good performance in separating the symmetrical and unsymmetrical modes without significant cross-coupling problems.

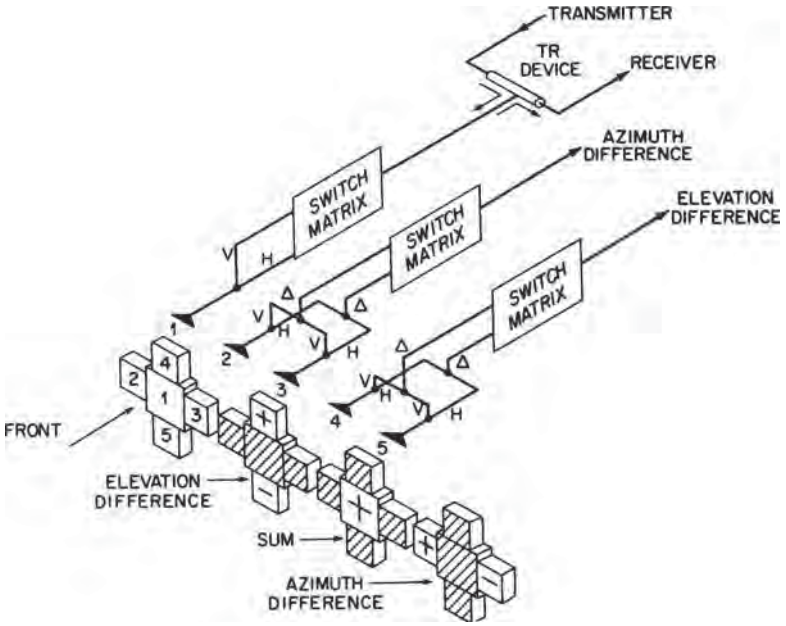


FIGURE 9.8 Five-horn feed with coupling to both linear-polarization components, which are combined by the switch matrix to select horizontal, vertical, or circular polarization

Multiband monopulse feed configurations are practical and in use in several systems. A simple example is a combined X-band and K_a-band monopulse paraboloid antenna radar. Separate conventional feeds are used for each band, with the K_a-band feed as a Cassegrain feed and the X-band feed at the focal point.[13] The Cassegrain subdish is a hyperbolic-shaped highly efficient grid of wires reflective to parallel polarization and transparent to orthogonal polarization. It is oriented to be transparent to the X-band focal-point feed behind it and reflective to the orthogonally polarized K_a-band feed extending from the vertex of the paraboloid.

Monopulse feed horns at different microwave frequencies can also be combined with concentric feed horns. The multiband feed clusters will sacrifice efficiency but can satisfy multiband requirements in a single antenna.

AGC (Automatic Gain Control). To maintain a stable closed-loop servosystem for angle tracking, the radar must maintain essentially constant loop gain independent of target echo size and range. The problem is that monopulse difference signals from the antenna are proportional to both the angle displacement of the target from the antenna axis and the echo signal amplitude. For a given tracking error, the error voltage would change with echo amplitude and target range causing a corresponding change in loop gain.

AGC is used to remove the angle-error-detector-output dependence on echo amplitude and retain constant tracking loop gain. A typical AGC technique is illustrated in Figure 9.9 for a one-angle coordinate tracking system. The AGC system detects the peak voltage of the sum signal and provides a negative dc voltage proportional to the peak signal voltage. The negative voltage is fed to the IF amplifier stage, where it is used to decrease gain as the signal increases. A high gain in the AGC loop is equivalent to dividing the IF output by a factor proportional to its amplitude.

In a three-channel monopulse radar, all three channels are controlled by the AGC voltage, which effectively performs a division by the magnitude of the sum signal or echo amplitude. Conventional AGC essentially holds constant gain during the pulse repetition interval. Also, the AGC of the sum channel normalizes the sum echo pulse amplitude to similarly maintain a stable range-tracking servo loop.

The angle-error detector, assumed to be a produce detector, has an output

$$|e| = k\frac{\Delta\Sigma}{|\Sigma||\Sigma|}\cos\theta \tag{9.2}$$

where $|e|$ is the magnitude of the angle-error voltage. Phases are adjusted to provide 0° or 180° on a point-source target. The resultant is

$$|e| = \pm k\frac{\Delta}{|\Sigma|} \tag{9.3}$$

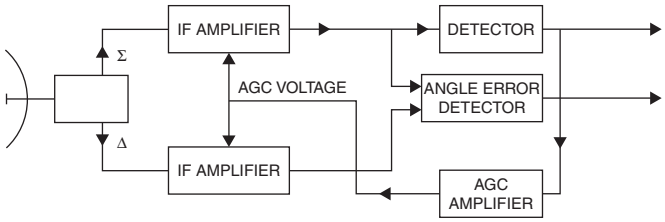


FIGURE 9.9 AGC in monopulse tracking

Complex targets can cause other phase relations as a part of the angle scintillation phenomenon.[1] The above error voltage, proportional to the ratio of the difference signal divided by the sum signal, is the desired angle-error-detector output, giving a constant angle error sensitivity.[1]

With limited AGC bandwidth, some rapid signal fluctuations modulate $|e|$ but the long-time-average angle sensitivity is constant. These fluctuations are largely from rapid changes in target reflectivity, $\sigma(t)$, that are from target amplitude scintillation. The random modulation of $|e|$ causes an additional angle noise component that affects the choice of AGC bandwidth.

The AGC performance in conical-scan radars provides similar constant angle error sensitivity. One major limitation in conical-scan radars is that the AGC bandwidth must be sufficiently lower than the scan frequency to prevent the AGC from removing the modulation containing the angle error information.

Phase-Comparison Monopulse. A second monopulse technique is the use of multiple antennas with overlapping (nonsquinted) beams pointed at the target. Interpolating target angles within the beam is accomplished, as shown in Figure 9.10, by comparing the phase of the signals from the antennas (for simplicity a single-coordinate tracker is described). If the target were on the antenna boresight axis, the outputs of each

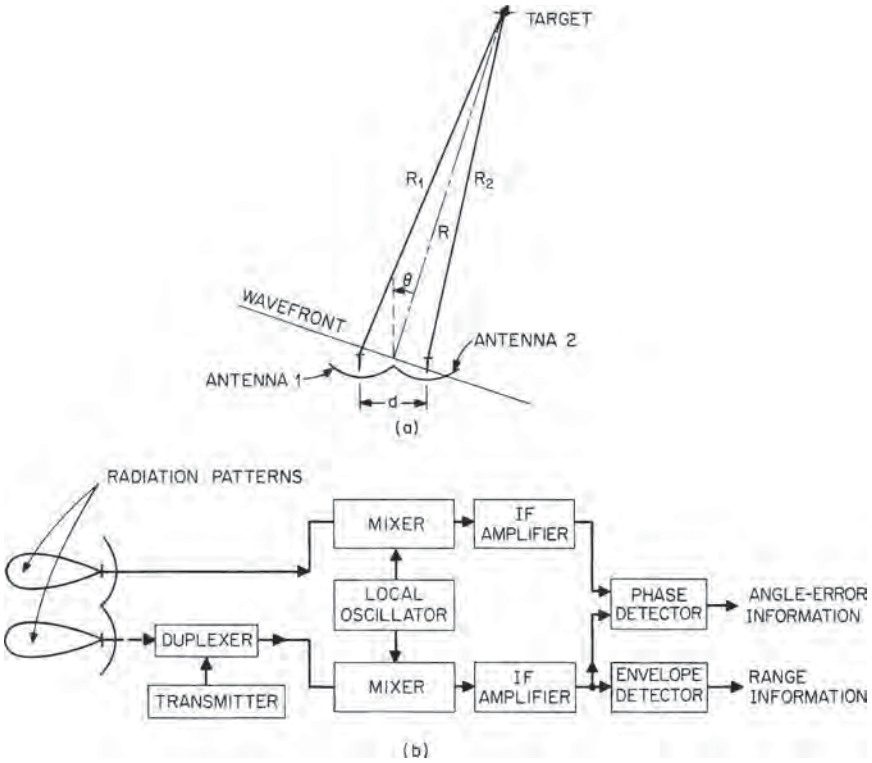


FIGURE 9.10 (*a*) Wavefront phase relationships in a phase comparison monopulse radar and (*b*) block diagram of a phase comparison monopulse radar (one angle coordinate)

individual aperture would be in phase. As the target moves off axis in either direction, there is a change in relative phase. The amplitudes of the signals in each aperture are the same so that the output of the angle-error phase detector is determined by the relative phase (see Figure 9.11). The phase-detector circuit is adjusted with a 90° phase shift on one channel to give zero output when the target is on axis and an output increasing with increasing angular displacement of the target with a polarity corresponding to the direction of error.

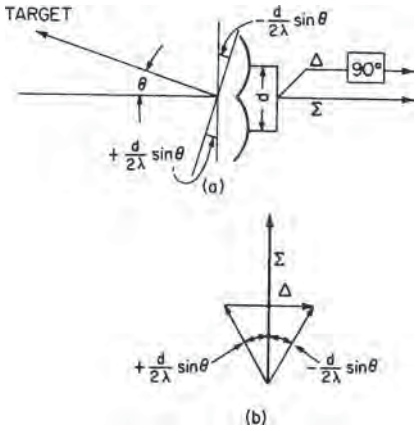


FIGURE 9.11 (*a*) RF phase-comparison monopulse system with sum and difference outputs and (*b*) vector diagram of the sum and difference signals

Typical flat-face corporate-fed phased arrays compare the output of halves of the aperture and fall into the class of phase-comparison monopulse. However, the basic signal processing of amplitude- and phase-comparison monopulse is similar, but the control of amplitude distribution across an array aperture for the sum and difference signals maintains efficiency and lower sidelobes.

Figure 9.10 shows the antenna and receiver for one angular-coordinate tracking by phase comparison monopulse. Any phase shifts occurring in the mixer and IF amplifier stages causes a shift in the boresight of the system. The disadvantages of phase-comparison monopulse with separate apertures compared with amplitude-comparison monopulse are the relative difficulty in maintaining a highly stable boresight and the difficulty in providing the desired antenna illumination taper for both sum and difference signals. The longer paths from the antenna outputs to the comparator circuitry make the phase-comparison system more susceptible to boresight change due to mechanical loading (sag), differential heating, etc.

A technique giving greater boresight stability combines the two antenna outputs at RF with passive circuitry to yield sum and difference signals, as shown in Figure 9.11. These signals may then be processed like a conventional amplitude-comparison monopulse receiver. The system shown in Figure 9.11 would provide a relatively good difference-channel taper, having smoothly tapered *E*-fields on each antenna. However, a sum-signal excitation with the two antennas provides a two-humped in-phase *E*-field distribution that causes high sidelobes since it looks like a two-element array. This problem may be reduced by allowing some aperture overlap but at the price of loss of angle sensitivity and antenna gain.

Electronic Scan Phased Array Monopulse. Tracking radars dedicated to single target tracking can provide very high precision long range performance, such as the AN/FPQ-6[14,15,16] (Figure 9.1*a*) with a specified precision of 0.05 milliradian. With high power and a high gain antenna (52 dB) and special tracking techniques, they are the workhorse for precision tracking of satellites and similar tasks. However, most modern tasks require precision simultaneous tracking of multiple simultaneous targets where use of multiple single target tracking radars are not cost effective. The development of electronic scan phased array technology has resulted in versatile high precision monopulse tracking with the capability of simultaneous multitarget tracking by switching its beam to each of several targets on a pulse-to-pulse basis or by groups

of pulses. Monopulse tracking is necessary to obtain angle data on each pulse to maintain adequate data rates when sharing pulses and power among several targets. A detailed discussion of electronic scan phased arrays is given in Chapter 13; however, some characteristics of the arrays require special consideration for the angle tracking performance of tracking radars using monopulse phased array antennas.

Optical-feed Monopulse Electronic Scan Arrays. Optical-feed monopulse arrays include the lens array and reflectarray (Chapter 13) that are optically fed by a conventional monopulse feed. The AN/MPQ-39[17] (Figure 9.1*b*) is an example of an optically fed array lens with the antenna mounted on a two-axis pedestal. Typical instantaneous electronic angle coverage is ±45° to an almost ±60° cone field-of-view that may be moved by pedestal drive to center on a multitarget event or follow an event progressing to a different area. Some military systems such as the Patriot with the ±60° cone of instantaneous view is fixed on its vehicle without a pedestal and is dependent on movement of its vehicle to change the region of angular coverage as needed. The advantages of space fed arrays are

- Conventional monopulse microwave horn feeds are used.
- Array elements are available with selectable polarization of the radiated energy when fed by an optimized linear polarized monopulse feed (such as in Figure 9.7) and selectable receive-polarization as well. This avoids the typical compromise and greater complexity of a polarization-controlled monopulse feed as described in Figure 9.8.
- Electronic scan array lenses can also refocus from a transmit feed horn to an adjacent receive feed horn on reception to allow high power transmission through a simple single horn feed to simplify isolation of the receiver from the transmit power.
- Arrays allow greater flexibility to optimize amplitude distribution of the radiated energy across the array to reduce sidelobes.

Most of the electronic scan phased array disadvantages are described in Chapter 13 and include losses in the array phase shifting elements, limitation of instantaneous bandwidth with conventional phase control elements (improved with special true-time-delay phase shifting), phase quantization errors (Chapter 13) resulting from phase shifting in steps, restriction to a single rf band (multiband arrays require special techniques with major compromises), and gradual degradation of performance as the beam is scanned from the normal to the array. The quantization errors from phase shifting in steps are of concern to monopulse radar because it results in corresponding random error steps in the electronic axis of the array. As described in Chapter 13, the quantization errors are inversely proportional to the number of phase shifting elements and 2^P where P is the number of bits of phase control in each element. Consequently, the high precision tracking radars with typically 4000 to 8000 phase shifters and four or more phase shift bits have small resultant electrical axis error steps on the order of 0.1 milliradians or less. The electrical axis errors are essentially random and can be further reduced by averaging. Intentional dither of phase steps may be introduced to aid in averaging.

The optically fed technique results in feed energy spillover around the aperture; however, these resultant spillover sidelobes can be eliminated by an absorbing cone between the feed and the array aperture. The absorbing cone is observed in the AN/MPQ-39 (Figure 9.1*b*). However, cooling is also necessary and provided, as observed by the cooling coils around the absorbing cone.

Of further concern to high precision monopulse applications is drift of the electronic axis that causes variations in phase and temperature variation across the array surface that causes distortion of the lens. Significant variation of heat distribution across the array face can result from high power transmitted through the phase shifting elements as well as the electronic phase control. Consequently, where high precision tracking is required, special cooling techniques may be necessary to maintain constant temperature across the aperture.

Corporate Feed Monopulse Electronic Scan Phased Array. The corporate feed array is fed by dividing and subdividing the transmit signal through transmission lines typically to subarrays of multiple array radiating elements. This technique, although typically resulting in heavier and higher cost implementation, offers the advantage of flexibility of control of the signal paths through the array structure, as described in Chapter 13. Another advantage is the capability to transmit very high peak power without the limitations of full peak power propagating through a single transmission line. This is accomplished in the corporate feed array by placing high power amplifiers where the power divides to the subarrays, allowing the sum of the high peak power amplifier outputs to add in space to meet requirements for long-range tracking and power sharing between multiple simultaneous targets.

The parallel power amplifier configuration also provides a practical means for overcoming the narrow instantaneous bandwidth of typical phased arrays at wide scan angles. Full array instantaneous bandwidth requires equal path lengths between each array element and the target, requiring many wavelengths of phase control or the equivalent time delay in array elements at wide angle scans. However, this control has prohibitively high loss for typical phased array radiating elements; consequently, typical phased array elements provide only sufficient phase control of up to 360° or to one wavelength, limited to tolerable loss, to cause the signal from each element to arrive approximately in phase at the target. Unfortunately, this shortcut is adequate for only a narrow instantaneous bandwidth. The parallel power amplifiers, as described above, provide a low power amplifier drive stage where the high loss of the desired time delay control can be tolerated to gain wide instantaneous bandwidth, as described in Chapter 13. The time delay may be controlled similar to the diode phase shifters used in radiating elements that switch between different line lengths to adjust phase. Longer time delay transmission line could be similarly controlled by diode switching to provide the wide instantaneous bandwidth to allow, for example, use of wideband narrow pulses to provide the range resolution requirements for tracking radar applications.

Two-Channel Monopulse. Monopulse radars may be designed with fewer than the conventional three IF channels. This is accomplished, for example, by combining the sum and difference signals in two IF channels and the sum and two difference signal outputs are then individually retrieved at the output. These techniques provide some advantages in AGC or other processing techniques but at the cost of reduced SNR, reduced angle data rate, and potential for cross coupling between azimuth and elevation information.

A two-channel monopulse receiver[18] combines the sum and difference signals at RF, as shown in Figure 9.12. The microwave resolver is a mechanically rotated RF coupling loop in cylindrical waveguide. The azimuth and elevation difference signals are excited in this guide with *E*-field polarization oriented at 90°. The energy in the coupler contains both difference signals coupled as the cosine and sine of the angular position of the coupler, $\omega_s t$, where ω_s is the angular rate of rotation. The hybrid adds the combined difference signals Δ at the angular

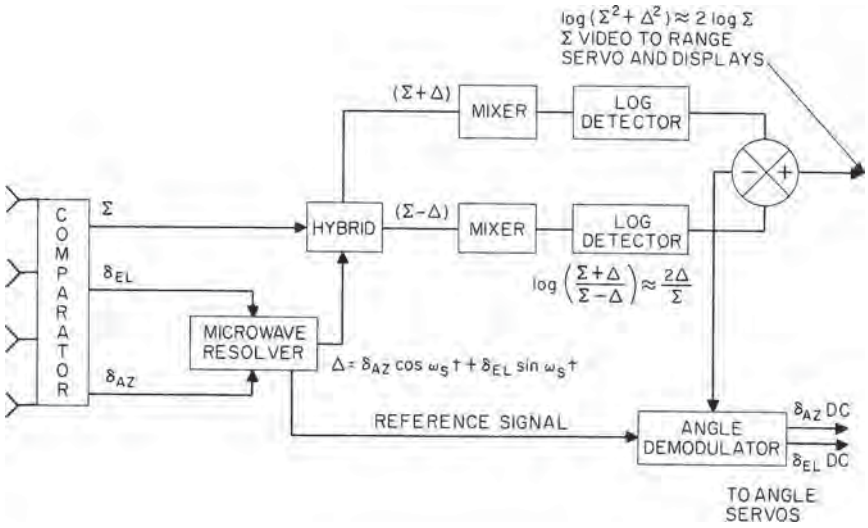


FIGURE 9.12 Block diagram of a two-channel monopulse radar system (*from R. S. Noblit*[18])

rate of rotation. The $\Sigma + \Delta$ and $\Sigma - \Delta$ outputs each look like the output of a conical-scan tracker except that their modulation function differs by 180°. In case one channel fails, the radar can be operated as a scan-on-receive-only conical-scan radar with essentially the same performance as a conical-scan radar. The advantage of two channels with opposite-sense angle-error information on one channel with respect to the other is that signal amplitude fluctuations in the received signal are canceled in the post-detection subtraction at the IF output that retrieves the angle-error information. The log IF performs essentially as an instantaneous AGC, giving the desired constant angle-error sensitivity of the difference signals normalized by the sum signal. The detected Δ information is a bipolar video where the error information is contained in the sinusoidal envelope. This signal is separated into its two components, azimuth- and elevation-error information, by an angle demodulation. The demodulator, using a reference from the drive on the rotating coupler, extracts the sine and cosine components from Δ to give the azimuth- and elevation-error signals. The modulation caused by the microwave resolver is of concern in instrumentation radar applications because it adds spectral components in the signal, complicating the possible addition of pulse doppler tracking capability to the radar.

This system provides instantaneous AGC operation with only two IF channels and operation with reduced performance in case of failure of either channel. However, there is a loss of 3-dB SNR at the receiver inputs, although this loss is partly regained by coherent addition of the Σ-signal information. The design of the microwave resolver must minimize loss through the device, and precisely matched IF channels are required to minimize cross coupling between the azimuth and elevation channels. In some modern systems, the resolver performance is improved by use of ferrite switching devices to replace the mechanical rotating coupler.

Conopulse. Conopulse (also called *scan with compensation*) is a radar tracking technique that is a combination of monopulse and conical scan.[19,20] A pair of antenna

beams is squinted in opposite directions from the antenna axis and rotated like a pair of conical-scan-radar beams. Since they exist simultaneously, monopulse information can be obtained from the pair of beams. The plane in which monopulse information is measured rotates. Consequently, elevation and azimuth information is sequential and must be separated for use in each tracking coordinate. Conopulse provides the monopulse advantage of avoiding errors caused by amplitude scintillation, and it requires only two receivers. However, it has the disadvantage of lower angle data rates and the mechanical complexity of providing and coupling to a pair of rotating antenna feedhorns.

9.3 SCANNING AND LOBING

The first technique used for radar angle tracking was to displace the antenna beam above and below the target in elevation and side to side of the target in azimuth to compare beam amplitudes similar to monopulse radar simultaneous lobing but differing by being in a time sequence. This was performed by a continuous conical beam scan, as illustrated in Figure 9.13[21] or by sequentially lobing up/down and right/left and observing the difference between amplitudes as a measure of displacement of the antenna axis from the target. The signal output for a conical-scan radar, illustrated in Figure 9.14, is typically a sinusoid amplitude modulation of the received target echo pulses. The amplitude of the modulation is a measure of the magnitude of the angle error, and the phase, relative to the scanning-beam rotation angle, indicates the portion of the error caused by each tracking axis.

The performance of scanning and lobing radar relative to the beam offset angle is described in Barton.[22] An optimum beam offset is described as a compromise between the loss of antenna gain and the increase in sensitivity to target angle displacement from the antenna axis as beam offset is increased. The optimum offset is typically chosen to provide the minimum rms angle-tracking error as affected by the signal-to-noise ratio and tracking sensitivity. Special tracking radar applications with nontypical requirements could arrive at a different optimum beam offset.

A major limitation of scanning and lobing radar is the susceptibility to target amplitude fluctuations that occur during the time the beam is moved from side to side or up and down. It is also susceptible to false modulation on signals from countermeasures. The echo fluctuations not related to antenna beam position cause false target angle-tracking errors.

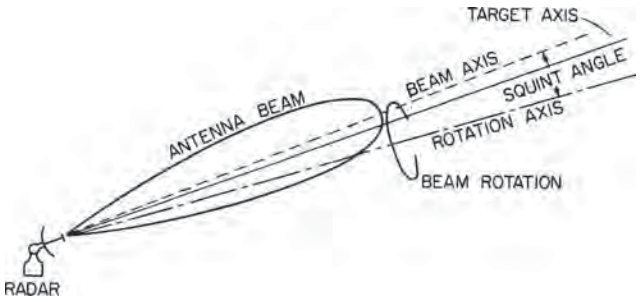


FIGURE 9.13 Conical-scan tracking

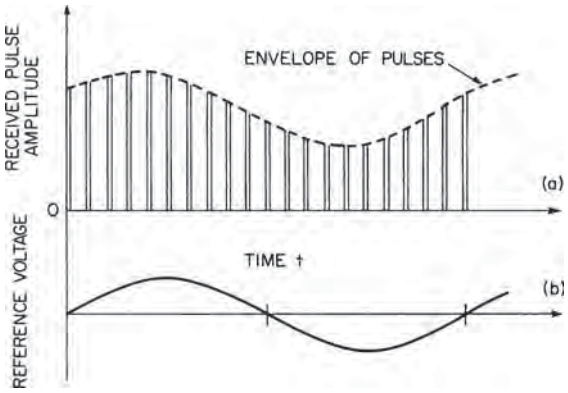


FIGURE 9.14 (*a*) Angle error information contained in the envelope of the received pulses in a conical-scan radar and (*b*) reference signal derived from the drive of the conical-scan feed

Monopulse radar was developed to provide simultaneous offset antenna beams for comparison of target echo amplitudes on a single pulse independent of echo signal amplitude fluctuations. However, few microwave devices and components were initially available, and the first monopulse systems were complex and resulted in cumbersome and inefficient antennas. At present, modern monopulse radars, as described in Section 9.2, provide highly stable and efficient antennas with high precision performance and have generally displaced scanning and lobing tracking radars for meeting the increasing demands for high precision and high data rate of angle information on each pulse. However, special radar tracking requirements may exist where a practical implementation of conical scan or lobing tracking radar may more effectively provide adequate performance.

9.4 SERVOSYSTEMS FOR TRACKING RADAR

The servosystem of a tracking radar is the subsystem of the radar that receives as its input the tracking-error voltage and performs the task of moving the antenna beam in a direction that will reduce to zero the alignment error between the antenna axis and the target. For two-axis tracking with a mechanical-type antenna pedestal, there are typically separate axes of rotation for azimuth and elevation and separate servosystems to move the antenna about each axis. A conventional servosystem is composed of amplifiers, filters, and a motor that moves the antenna in a direction to maintain the antenna axis on the target. Range tracking is accomplished by a similar system to maintain range gates centered on the received echo pulses. This may be accomplished by analog techniques or by digital-counter registers that retain numbers corresponding to target range to provide a closed range-tracking loop digitally.

Servosystems may use hydraulic-drive motors, conventional electric motors geared down to drive the antenna, or direct-drive electric motors where the antenna mechanical axis shaft is part of the armature, and the motor field is built into the support case. The direct drive is heavier for a given horsepower but eliminates

gear backlash. Backlash may also be reduced with conventional motors by duplicate parallel drives with a small residual opposing torque when near zero angle rate. Amplifier gain and filter characteristics as well as motor torque and inertia determine the velocity and acceleration capability or the ability to follow the higher-order motion of the target.

It is desired that the antenna beam follow the center of the target as closely as possible, which implies that the servosystem should be capable of moving the antenna quickly. The combined velocity and acceleration characteristics of a servosystem can be described by the frequency response of the tracking loop, which acts essentially like a low-pass filter. Increasing the bandwidth increases the quickness of the servosystem and its ability to follow a strong, steady signal closely. However, a typical target causes scintillation of the echo signal, giving erroneous error-detector outputs, and at long range, the echo is weak, allowing receiver noise to cause additional random fluctuations on the error detector output. Consequently, a wide servo bandwidth, which reduces lag errors, allows the noise to cause greater erroneous motions of the tracking system. Therefore, for best overall performance, it is necessary to limit the servo bandwidth to the minimum necessary to maintain a reasonably small tracking-lag error. There is an optimum bandwidth that may be chosen to minimize the amplitude of the total erroneous outputs including both tracking lag and random noise, depending upon the target, its trajectory, and other radar parameters.

The optimum bandwidth for angle tracking is range-dependent. A target with typical velocity at long range has low angle rates and a low SNR, and a narrower servo passband will follow the target with reasonably small tracking lag while minimizing the response to receiver thermal noise. At close range, the signal is strong, overriding receiver noise, but target angle scintillation errors proportional to the angular span of the target are large. A wider servo bandwidth is needed at close range to keep tracking lag within reasonable values, but it must not be wider than necessary or the target angle scintillation errors, which increase inversely proportional to target range, may become excessive.

The low-pass closed-loop characteristic of a servosystem is unity at zero frequency, typically remaining near this value up to a frequency near the low-pass cutoff, where it may peak up to higher gain, as shown in Figure 9.15*a*. The peaking is an indication of system instability but is allowed to be as high as tolerable, typically to about 3 dB above unity gain to obtain maximum bandwidth for a given servomotor drive system. System A in Figure 9.15*a* is a case of excessive peaking of about 8 dB. The effect of the peaking is observed by applying a step error input to the servosystem. The peaking of the low-pass characteristic results in an overshoot when the antenna axis moves to align with the target. High peaking causes a large overshoot and a return to the target with additional overshoot. In the extreme (as in system A shown in Figure 9.15*b*), the antenna zeros in on the target with a damped oscillation. An optimum system compromise between speed of response and overshoot, as in system B, allows the antenna to make a small overshoot with reasonably rapid exponential movement back to the target. This corresponds to about 1.4 dB peaking of the closed-loop low-pass characteristic.

The resonant frequency of the antenna and servosystem structure (including the structure foundation, which is a critical item) must be kept well above the bandwidth of the servosystem, otherwise the system can oscillate at the resonant frequency. A factor of at least 10 is desirable for the ratio of system resonance frequency to servo bandwidth. High resonant frequency is difficult to obtain with a large antenna, such as the AN/FPQ-6 radar with a 29-ft dish, because of the large mass of the system.

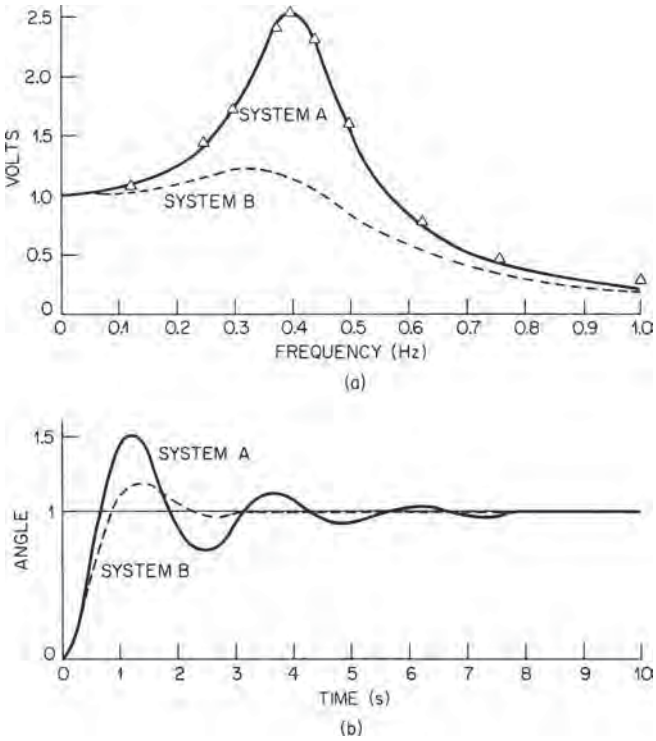


FIGURE 9.15 (*a*) Closed-loop frequency-response characteristics of two servosystems and (*b*) their corresponding time response to a step input

The ratio was pushed to a very minimum of about 3 to obtain servosystem bandwidth of the specified 3.5 Hz. A smaller radar with a 12-ft dish, for example, can provide a servosystem bandwidth up to 7 or 8 Hz with conventional design.

Locke[23] describes methods for calculating angle tracking lag for a given target trajectory versus time and set of servosystem characteristics. Range tracking lags may be similarly calculated, but with typical inertialess electronic tracking systems, tracking lags are usually negligible.

Electronically steerable arrays provide a means for inertialess angle tracking. However, because of this capability, the system can track multiple targets by rapidly switching from one to another rather than continuously tracking a single target.

The tracker simply places its beam at the location where the target is expected, corrects for the pointing error by converting error voltages (with known angle-error sensitivity) to units of angle, and moves to the next target. The system determines where the target was and, from calculations of target velocity and acceleration, predicts where it should be the next time the beam looks at the target. The lag error, in this case, is dependent on many factors, including the accuracy of the value of angle sensitivity used to convert error voltages to angular error, the size of the previous tracking error, and the time interval between looks.

9.5 TARGET ACQUISITION AND RANGE TRACKING

Range tracking is accomplished by continuously measuring the time delay between the transmission of an RF pulse and the echo signal returned from the target, and converting the roundtrip delay to units of distance. The range measurement is the most precise position-coordinate measurement of the radar; typically, with high SNR, it can be within a few meters at hundreds-of-miles range. Range tracking usually provides the major means for discriminating the desired target from other targets (although doppler frequency and angle discrimination are also used) by performing range gating (time gating) to eliminate the echo of other targets at different ranges from the error-detector outputs. The range-tracking circuitry is also used for acquiring a desired target. Range tracking requires not only that the time of travel of the pulse to and from the target be measured but also that the return is identified as a target rather than noise and a range-time history of the target be maintained.

Although this discussion is for typical pulse-type tracking radars, range measurement may also be performed with CW radars using FM-CW, a frequency-modulated CW that is typically a linear-ramp FM. The target range is determined by the range-related frequency difference between the echo-frequency ramp and the frequency of the ramp being transmitted. The performance of FM-CW systems, with consideration of the doppler effect, is described in Sherman.[1]

Acquisition. The first function of the range tracker is acquisition of a desired target. Although this is not a tracking operation, it is a necessary first step before range tracking or angle tracking may take place in a typical radar. Some knowledge of target angular location is necessary for pencil-beam tracking radars to point their typically narrow antenna beams in the direction of the target. This information, called *designation data,* may be provided by surveillance radar or some other source. It may be sufficiently accurate to place the pencil beam on the target, or it may require the tracker to scan a larger region of uncertainty. The range-tracking portion of the radar has the advantage of seeing all targets within the beam from close range out to the maximum range of the radar. It typically breaks this range into small increments, each of which may be simultaneously examined for the presence of a target. When beam scanning is necessary, the range tracker examines the increments simultaneously for short periods, such as 0.1 s, makes its decision about the presence of a target, and allows the beam to move to a new location if no target is present. This process is typically continuous for mechanical-type trackers that move the beam slowly enough that a target will remain well within the beam for the short examination period of the range increments.

Target acquisition involves consideration of the *S/N* threshold and integration time needed to accomplish a given probability of detection with a given false-alarm rate similar to surveillance radar. However, high false-alarm rates, as compared with values used for surveillance radars, are used because the operator knows that the target is present, and operator fatigue from false alarms when waiting for a target is not involved. Optimum false-alarm rates are selected on the basis of performance of electronic circuits that observe each range interval to determine which interval has the target echo.

A typical technique is to set a voltage threshold sufficiently high to prevent most noise peaks from crossing the threshold but sufficiently low that a weak signal may cross. An observation is made after each transmitter pulse as to whether, in the range interval being examined, the threshold has been crossed. The integration time allows the radar to make this observation several times before deciding if there is a target present.

The major difference between noise and a target echo is that noise spikes exceeding the threshold are random, but if a target is present, the threshold crossings are more regular. One typical system simply counts the number of threshold crossings over the integration period, and if crossings occur for more than half the number of times that the radar has transmitted, a target is indicated as being present. If the radar pulse repetition frequency is 300 Hz and the integration time is 0.1 s, the radar will observe 30 threshold crossings if there is a strong and steady target. However, because the echo from a weak target combined with noise may not always cross the threshold, a limit may be set, such as 15 crossings, that must occur during the integration period for a decision that a target is present. For example, performance on a non-scintillating target has a 90% probability of detection at a 2.5 dB-per-pulse SNR and a false alarm probability of 10^{-5}. The AN/FPS-16 and AN/FPQ-6 instrumentation radars use these detection parameters with 10 contiguous range gates of 1000 yd each for acquisition. The 10 gates give coverage of a 5-nmi range interval at the range where the target is expected, possibly from coarse range designation from search radar.

Range Tracking. Once a target is acquired in range, it is desirable to follow the target in the range coordinate to provide distance information or slant range to the target. Appropriate timing pulses provide range gating so the angle-tracking circuits and AGC circuits look at only the short range interval, or time interval, when the desired echo pulse is expected. The range tracking operation is performed by closed-loop tracking similar to the angle tracker. Error in centering the range gate on the target echo pulse is sensed, error voltages are generated, and circuitry is provided to respond to the error voltage by causing the gate to move in a direction to recenter on the target echo pulse.

The range-tracking error may be sensed in many ways. The most commonly used method is the early- and late-gate technique (see Figure 9.16). These gates are timed so that the early gate opens at the beginning of the main range gate and closes at the center of the main gate. The late gate opens at the center and closes at the end of the main range gate. The early and late gates each allow the target video to charge capacitors during the time when the gates are open. The capacitors act as integrators. The early-gate capacitor charges to a voltage proportional to the area of the first half of the target video pulse, and the late-gate capacitor charges negatively proportionally to the late half of the target video. When the gates are properly centered about a symmetrical video pulse, the capacitors are equally charged. Summing their charge voltages yields a zero output.

When the gates are not centered about the target video, so that the early gate extends past the center of the target video, the early-gate capacitor charged positively receives a greater charge. The late gate sees only a small portion of the pulse, resulting in a smaller negative charge. Summing the capacitor voltages results in a negative output. Over a range of errors of approximately ± 1/4 of the target-video pulse width, the voltage output is essentially a linear function of timing error and of a polarity corresponding to the direction of error. During acquisition, the target is centered in the 1000 yd acquisition gate by range-tracking techniques described as follows, and the gate is reduced to approximately the width of the radar transmit pulse for normal tracking.

Many radar range-tracking systems use high speed sampling circuitry to take three to five samples in the vicinity of the echo video pulse. The amplitudes of the samples on the leading and lagging halves of the pulse are compared for range-error sensing similar to the comparison of amplitudes in the early-late-gates range tracker.

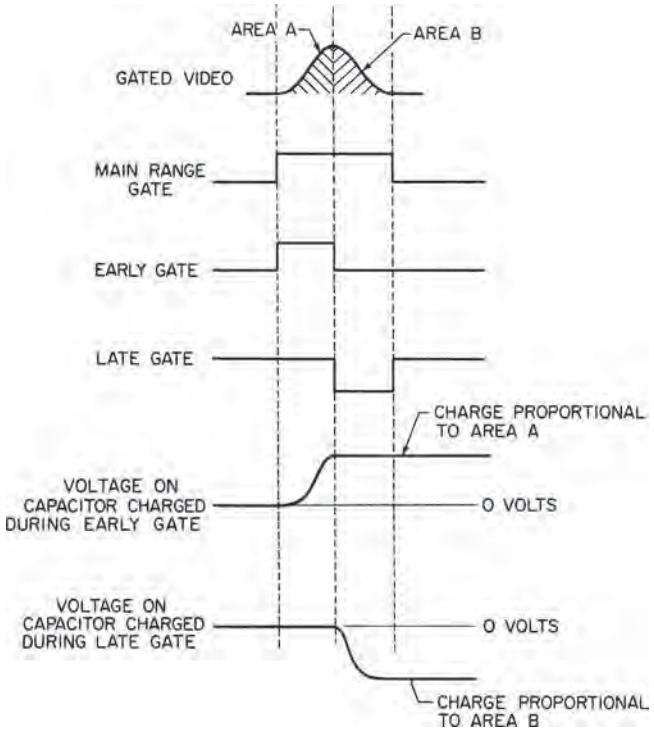


FIGURE 9.16 Early- and late-gate range-error-sensing circuit

In some cases, leading- or lagging-edge range tracking is desired. This has been accomplished in some applications by simply adding a bias to move the error-sensing gates either to lead or lag the center of the target. This provides some rejection by the gates of undesired returns that might occur near the target, such as the echoes from other nearby targets. Threshold devices are also used as leading- or lagging-edge trackers by observing when the target video exceeds a given threshold level. The point of crossing the threshold is used to trigger gating circuits to read out a target range from timing devices or to generate a synthetic target pulse.

The range-tracking loop is closed by using the range-error-detector output to reposition range gates and correct range readout. One technique uses a high-speed digital counter driven by a stable oscillator. The counter is reset to zero at the time of the transmit pulse. Target range is represented by a number stored in a digital register, as shown in Figure 9.17. A coincidence circuit senses when the digital counter reaches the number in the range register and generates the range gate, as indicated in the block diagram shown in Figure 9.18. A range error sensed by the range error detector results in an error voltage that drives a voltage-controlled variable-frequency oscillator to increase or decrease the count in the range register, depending on the polarity of the error voltage. This changes the number in the range register toward the value corresponding to the range of the target. Range readout is accomplished by reading the number in the register, where, for example, each bit may correspond to a 2-yd range step.

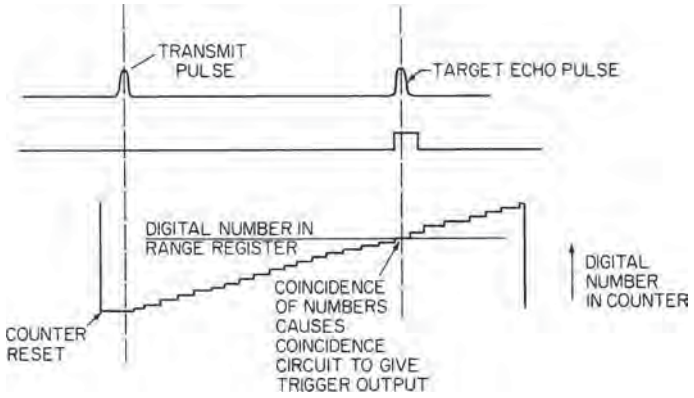

TRANSMIT PULSE
TARGET ECHO PULSE
DIGITAL NUMBER IN RANGE REGISTER
DIGITAL NUMBER IN COUNTER
COINCIDENCE OF NUMBERS CAUSES COINCIDENCE CIRCUIT TO GIVE TRIGGER OUTPUT
COUNTER RESET

FIGURE 9.17 Digital range tracker operation

Another technique is to use a pair of oscillators—one controlling the transmitter trigger and the other controlling the range gate.[24] The range rate is controlled by the beat frequency between the oscillators, where one is frequency-controlled by the range

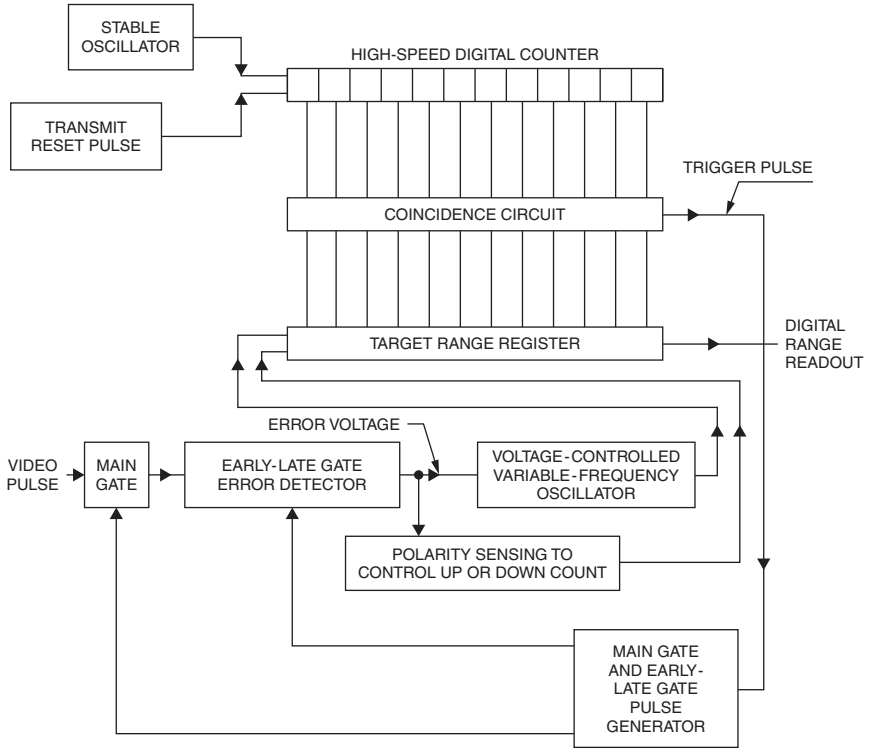


FIGURE 9.18 Block diagram of a digital range tracker

error-detector-output voltage. The beat frequency is a small fraction of one Hz and is better visualized as a phase rate between the transmit pulse cycle and cycle of the range gate. The changing phase causes the range gate to follow a moving target.

The electronic range tracker is inertialess, allowing any desired slew speed, and provides flexibility for conveniently generating acquisition gates for automatic-detection circuitry as well as transmitter trigger and pre-trigger pulses. Tracking bandwidth is usually limited to that necessary for tracking to minimize loss of track to false targets and countermeasures. Many other electronic range-tracking techniques also offering most of these advantages are used.[22]

***n*th-Time-Around Tracking.** To extend unambiguous range by reducing the PRF increases the acquisition time and reduces the data rate. A solution to this problem is called *nth-time-around tracking,* which avoids transmitting at the time that an echo is expected to arrive and can resolve the range ambiguity. This allows the radar to operate at high PRF and track unambiguously to long ranges where several pulses may be propagating in space to and from the target. The technique is useful only when a target is being tracked. During acquisition, the radar must look at the region between transmitter pulses, and upon initial acquisition, it closes the range- and angle-tracking loops without resolving the range ambiguity. The next step is to find which range interval, or between which pair of transmit pulses, the target is located. The zone *n* is determined by coding a transmit pulse and counting how many pulses return before the coded pulse returns.

Instrumentation radars provide *n*th-time-around tracking capability because beacons are used on rockets and space vehicles to provide sufficient signal level at very long ranges.

To prevent the target echo from being blanked by a transmit pulse, it is necessary to sense when the target is approaching an interference region and shift the region. This is accomplished by changing the PRF or alternately delaying groups of pulses equal to the number of pulses in propagation. This can be performed automatically to provide an optimum PRF shift or to alternately delay pulse groups of the correct number of pulses.

9.6 SPECIAL MONOPULSE TECHNIQUES

Dual-Band Monopulse. Dual-band monopulse can be efficiently accommodated on a single antenna to combine the complementary features of two RF bands.[13,25] A useful combination of bands is X band (9 GHz) and K_a band (35 GHz). The X-band operation provides the expected microwave performance of good radar range and precise tracking. Its weakness is the low-angle multipath region and the availability of electronic countermeasures in the band. The K_a band, although atmospheric- and rain-attenuation-limited, provides much greater tracking precision in the low-angle multipath region and a second and more difficult band that the electronic-countermeasures techniques must cover.

A Naval Research Laboratory system called TRAKX (Tracking Radar At K_a and X bands) was designed for instrumentation radar applications for missile and training ranges.[13] Its purpose was to add precision tracking on targets, essentially to "splash" and provide precision tracking at K_a band in an environment of X-band countermeasure experiments.

A similar X- and K_a-band system was developed by Hollandse Signaalapparaten of the Netherlands for tactical application. The land-based version called FLY-CATCHER is part of a mobile anti-air-warfare system.[25] Another version, GOAL-KEEPER, is for a shipboard anti-air-warfare application for the fire control of Gatling guns.[26] Both systems take full advantage of the two bands to provide precision tracking in multipath and electronic-countermeasures environments.

Mirror-Scanned Antenna (Inverse Cassegrain). An antenna technique that uses a movable RF mirror for scanning the beam, called a *mirror-scanned antenna* or *inverse Cassegrain,* provides useful applications to monopulse radar. The technique uses a radome-supported wire-grid paraboloid that reflects parallel-polarized feed energy. The beam, polarized parallel to the grid, is collimated by the paraboloid and is reflected by a flat moveable polarization rotating mirror. The basic polarization rotating mirror is a flat metal surface with a grid of wires located a quarter wavelength above the metal surface and oriented at 45° relative to the RF energy reflected from the paraboloid. The RF energy may be visualized as being composed of a component parallel to and reflecting from the grid and a component perpendicular to and passing through the grid to reflect from the metal mirror surface below. By traveling the quarter wavespace twice, this component is shifted by 180° in phase. When added to the reflection from the grid, it results in a 90° change in polarization. The total reflected energy from the mirror rotated by 90° will efficiently pass through the wire-grid paraboloid. The advantages are as follows: (1) The mirror and its drive mechanism are the only moving parts for beam movement. The feed and radome-supported paraboloid remain fixed. (2) The beam movement is by specular reflection, twice the angle of the mirror tilt. This provides a compact structure for a given angle coverage requirement. (3) The normally lightweight mirror and the 2:1 beam displacement versus mirror tilt allow reduced size and very rapid beam scan with low servo drive power.

The compactness and lightness are particularly attractive for airborne applications such as the Thompson-CSF Agave radar in the Super Entendards, which determines target range and designation data for the Exocet missile. It is a compact monopulse roll- and pitch-stabilized radar with 140° azimuth and 60° elevation scan.[27] The Israeli Elta subsidiary of Israeli Aircraft Industries also developed an airborne tracking radar using this antenna technology for air-to-air combat and ground weapon delivery.[28]

A ground- or shipboard-based experimental mirror antenna system concept was developed with dual-band monopulse capability (3.0 GHz and 9.3 GHz bands). The objective included high-speed beam movement for high-data-rate 3D surveillance and multitarget precision tracking.[29] Dual-band polarization-twist mirror design was accomplished with a two-layer mirror grid configuration.[30]

On-Axis Tracking. The best radar tracking performance is usually accomplished when the target is essentially on the radar antenna axis. Therefore, for maximum precision tracking, it is desirable to minimize lag and other error sources affecting the beam pointing. A technique called *on-axis tracking* was developed to minimize radar axis deviation from the target by prediction and optimum filtering within the tracking loop.[8,31] The technique is particularly effective when the target trajectory is known approximately, such as when tracking satellites in orbit or a ballistic target. A computer in the tracking loop can cause the radar to follow an estimated set of orbital parameters, for example. It also performs optimum filtering of radar angle-error-detector output to generate an error trend from which it can update the assumed set

of orbital parameters to correct the radar beam movement to update the original set of orbital parameters, and by this means, the radar antenna axis can be held on target with minimum error.

Improved tracking can also be provided on other targets where the approximate trajectory can be anticipated. However, performance of on-axis tracking is limited when tracking targets with unanticipated maneuvers.

9.7 SOURCES OF ERROR

There are many sources of error in radar-tracking performance. Fortunately, most are insignificant except for very high-precision tracking-radar applications such as range instrumentation, where the angle precision required may be of the order of 0.05 mrad (mrad, or milliradian, is one thousandth of a radial, or the angle subtended by 1-m cross-range at 1000-m range). Many sources of error can be avoided or reduced by radar design or modification of the tracking geometry. Cost is a major factor in providing high-precision-tracking capability. Therefore, it is important to know how much error can be tolerated, which sources of error affect the application, and what is the most cost-effective means to satisfy the accuracy requirements.

Because tracking radars track targets not only in angle but also in range and sometimes in doppler, the errors in each of these target parameters must be considered on most error budgets. The rest of this chapter will provide a guide for determining the significant error sources and their magnitudes.

It is important to recognize what the actual radar information output is. For a mechanically moved antenna, the angle-tracking output is usually obtained from the shaft position of the elevation and azimuth antenna axes. Absolute target location (relative to earth coordinates) will include the accuracy of the survey of the antenna pedestal site.

Phased array instrumentation radar, such as the Multi-object Tracking Radar (MOTR), provide electronic beam movement over a limited sector of about $\pm 45^{o}$ to approximately $\pm 60^{o}$ plus mechanical movement of the antenna to move the coverage sector.[16–19] The output will be mechanical shaft positions locating the normal to the array plus digital angle information from the electronic beam scan for each target.

9.8 TARGET-CAUSED ERRORS (TARGET NOISE)

Radar tracking of targets is performed by use of the echo signal reflected from a target illuminated by the radar transmit pulse. This is called *skin tracking* to differentiate it from *beacon tracking,* where a beacon or a transponder transmits its signal to the radar and usually provides a stronger point-source signal. Because most targets, such as aircraft, are complex in shape, the total echo signal is composed of the vector sum of a group of superimposed echo signals from the individual parts of the target, such as the engines, propellers, fuselage, and wing edges. The motions of a target with respect to the radar causes the total echo signal to change with time, resulting in random fluctuations in the radar measurements of the parameters of the target. These fluctuations caused by the target only, excluding atmospheric effects and radar receiver noise contributions, are called *target noise.*

This discussion of target noise is based largely on aircraft, but it is generally applicable to any target, including land targets of complex shape that are large with respect to a wavelength. The major difference is in the target motion, but the discussions are sufficiently general to apply to any target situation.

The echo return from a complex target differs from that of a point source by the modulations that are produced by the change in amplitude and relative phase of the returns from the individual elements. The word *modulations* is used in plural form because five types of modulation of the echo signal that are caused by a complex target affect radars. These are amplitude modulation, phase front modulation (glint), polarization modulation, doppler modulation, and pulse time modulation (range glint). The basic mechanism by which the modulations are produced is the motion of the target, including yaw, pitch, and roll, which causes the change in relative range of the various individual elements with respect to the radar.

Although the target motions may appear small, a change in relative range of the parts of a target of only one-half wavelength (because of the two-way radar signal path) causes a full 360° change in relative phase. At X band, this is about 1.5 cm, which is small even compared with the flexure between parts of an aircraft.

The five types of modulation caused by a complex target are discussed next.

Amplitude Noise. Amplitude noise is the change in echo signal amplitude caused by a complex-shaped target, excluding the effects of changing target range. It is the most obvious of the various types of echo signal modulation by a complex-shaped target and is readily visualized as the fluctuating sum of many small vectors changing randomly in relative phase. Although it is called *noise,* it may include periodic components. Amplitude noise typically falls into a low frequency and high frequency region of interest. These categories overlap in some respects, but it is convenient to separate the noise in these two frequency ranges because they are generated by different phenomena, and they are each significant to different functions of the radar.

Low-Frequency Amplitude Noise. The low-frequency amplitude noise is the time variation of the vector sum of the echoes from all the reflecting surfaces of the target. The time variation is visualized by considering the target as a relatively rigid body with normal yaw, pitch, and roll motions. The small changes in relative range of the reflectors caused by this motion result in corresponding "random" change in the relative phases. Consequently, the vector sum fluctuates randomly. Typically, target random motion is limited to small aspect changes such that the amplitudes of the echoes from the individual reflectors vary little over a period of a few seconds, and change in relative phase is the major contributor. Exceptions are large flat surfaces with narrow reflection patterns.

An example of a target configuration is a distribution of reflecting surfaces that change in relative range with target motion. A typical pulse amplitude time function is a slowly varying echo amplitude.[32] The low-frequency amplitude noise contributes the largest portion of the noise modulation density and is concentrated mainly below about 10 Hz at X band. The amplitude-noise spectrum is similar for both large and small targets. This is because the rate of relative range change is a function of both angular yaw and distance from the center of gravity of the aircraft. Thus, a larger aircraft with slow yaw rates but greater wingspan generates a low-frequency noise spectrum similar to that of a small aircraft with high yaw rates but smaller wingspan. However, the larger aircraft typically has the broader noise spectrum because of the difference in distribution of dominant reflectors.

The radar frequency affects the low-frequency amplitude-noise spectrum shape where the spectrum width is closely proportional to the radar frequency (if the target span is assumed to be at least several wavelengths). The reason for this dependence is that the relative phase of the individual echo signals is a function of the number of wavelengths of change in relative range caused by the target's random motion. Thus, with shorter wavelengths, a given relative range change will subtend more wavelengths, causing higher phase rate, resulting in higher-frequency noise components. The rate of amplitude fluctuations of the envelope of the echo pulses is approximately proportional to the radar frequency.

A mathematical model of low-frequency amplitude noise of a typical aircraft is given by

$$A^2(f) = \frac{0.12B}{B^2 + f^2} \tag{9.4}$$

where $A^2(f)$ = (fractional modulation)^2/Hz
B = half-power bandwidth, Hz
f = frequency, Hz

The value of B falls typically between 1.0 Hz and 2.5 Hz at X band, with the larger aircraft at the higher values because of the larger reflectors, such as engines, spread out along the wings. These reflectors with the greater separation contribute to the higher frequencies because their relative range change is large for a given angular movement of the target. $A^2(f)$ is the modulation power density such that the spectrum may be integrated over any frequency range to find the total noise power within a frequency band of interest. Taking the square root of the value of the integral gives the rms modulation.

High-Frequency Amplitude Noise. High-frequency amplitude noise consists of both random noise and periodic modulation. The random noise is largely a result of the vibration and moving parts of the aircraft producing a relatively flat noise spectrum spread out to a few hundred Hz, depending on the type of aircraft. The rms noise density is typically a few percent of modulation per $\sqrt{\text{Hz}}$.

The periodic modulation appearing as spikes in the Figure 9.19 spectrum are caused by rapidly rotating parts of an aircraft, such as the propellers. As the echo from a propeller blade changes with aspect when it rotates, a periodic modulation

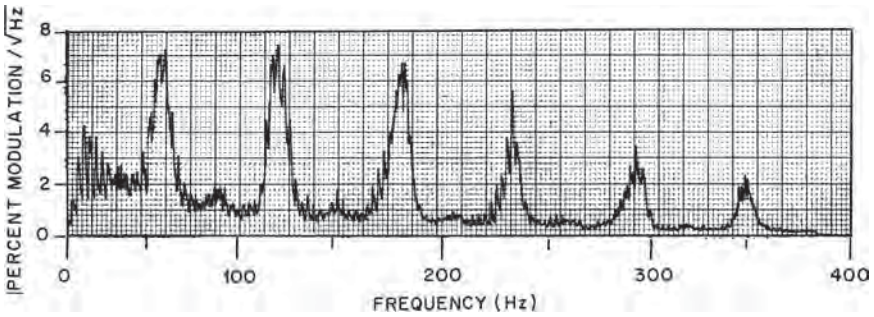


FIGURE 9.19 Typical amplitude spectral voltage distribution showing the propeller modulation measured on a propeller-driven aircraft in flight (*Figure 4 from Dunn, Howard, and King*[33] *© IRE 1959*)

is produced. The background noise from the airframe is also observed. The spikes in the spectrum result from a fundamental modulation frequency related to the propeller rev/min and number of blades. Since it is not usually sinusoidal, there are harmonic frequencies spread throughout the spectrum, as shown in Figure 9.19 for the SNB, a small aircraft with two propeller engines. The location of these spikes is not dependent on RF frequency, as in the case of low-frequency amplitude noise, because the target controls the periodicity of the modulation, which is dependent only on the aircraft propeller rotation rate and number of blades. Jet aircraft may also cause echo amplitude modulation of radar signals reflected from rotating fan blades from within the jet engines. The jet engine caused modulation is called Jet Engine Modulation (JEM) spectral modulation lines. The high-frequency-noise modulation affects scan-type tracking radars, as described later, and gives some information as to the type of aircraft.

Effects of Amplitude Scintillation on Radar Performance. Amplitude noise, to some extent, affects all types of radars in probability of detection and tracking radar accuracy.[32–36] One effect on all types of tracking radars is the interrelation between the low-frequency spectrum of amplitude noise, the AGC characteristics (which determine to what extent the slow fluctuations are smoothed), and the angle noise. The effects on angle noise are described later in this section, where it is described why a fast-acting AGC is generally the preferred choice for maximizing overall tracking accuracy.

High-frequency amplitude noise causes errors only in conical-scan or sequential lobing tracking radars because the effects are eliminated by the monopulse techniques. Conical scan or sequential lobing, to sense target direction, depend upon measuring the amplitude of the signal for at least two different antenna beam positions for each tracking axis. In azimuth tracking, for example, the antenna beam is displaced to the left of the target and then to the right. If the target were on the antenna axis, the signal would drop the same amount when the beam (assumed to be symmetrical) is moved an equal amount in either direction. The amplitudes for each beam position are subtracted in an angle error detector; hence, the output is zero if the target is on the antenna axis and becomes finite, increasing positively or negatively as the target moves off axis to the right or left.

High-frequency noise can cause the amplitude to change during the time taken to move the antenna beam from one position to the next. Even if the target is on axis, high-frequency noise can cause the amplitude at the two beam positions to differ, thus causing an erroneous indication that the target is off axis. This effect is averaged out except for the noise spectral energy near the scan rate. For example, a periodic modulation spike near the scan rate will cause the tracking radar to drive its antenna in a circular motion around the target at a rate equal to the difference in frequency between the scan rate and the frequency of the spectral line. The direction, clockwise or counterclockwise, depends upon whether the spectral line is above or below the scan rate and whether the scan is clockwise or counterclockwise. The servosystem filters out all frequencies outside the frequency range between the scan rate plus the servo bandwidth and the scan rate minus the servo bandwidth, and an angle sensitivity constant that converts rms modulation to rms angle error.

An equation using this relation to calculate rms noise in scanning and lobing-type tracking radars caused by high-frequency amplitude noise[22] is

$$\sigma_s = \frac{\theta_B}{k_s}\sqrt{A^2(f_s)\beta} \tag{9.5}$$

where σ_s = rms angle error in same angular units as θ_B
$A(f_s)$ = rms-fractional-modulation noise density in vicinity of scan rate
k_s = conical-scan error slope ($k_s = 1.6$ for system optimum[22])
θ_B = one-way antenna beamwidth
β = servo bandwidth, Hz

A sample calculation for an f_s of 120 Hz, where $A(f_s)$ from measured data taken on a large jet aircraft is approximately $0.018/\sqrt{\text{Hz}}$, θ_B is 25 mils, and β is 2 Hz, gives a σ_s of 0.42 mil rms.

In the case of a periodic modulation, where a spectral line falls within the band $f_s \pm \beta$, the rms noise is $\sigma_s = 0.67\theta_B A(f_s)$, where $A(f_s)$ is the rms fractional modulation caused by the spectral line. The resultant rms tracking error σ_s will be periodic at the frequency $f_s - f_t$ where f_t is the frequency of the spectral line.

The effects of amplitude noise on target detection and acquisition are of concern in all types of radars,[2] particularly at long ranges where the signal is weak. The amplitude fluctuations can cause the signal to drop below the noise level for short periods of time, thus affecting the choice of thresholds, acquisition scan rate, and detection logic.[34–36]

Angle Noise (Glint). Angle noise causes a change with time in the apparent location of the target with respect to a reference point on the target. This reference point is usually chosen as the *center of "gravity"* of the reflectivity distribution along the target coordinate of interest. The center of gravity is the long-time-averaged tracking angle on a target. The term *glint* is sometimes used for angle noise, but it gives the false impression that the wander in the apparent position of a target always falls within the target span. It was originally expected that angle fluctuations caused in a monopulse radar by a target would be simple variations in the center of gravity of the reflecting areas; however, much larger angle errors were observed. The apparent angular location of a target can fall at a point completely outside the extremities of the target. This can be demonstrated both experimentally and theoretically.[37,38] A pair of scatterers can be appropriately spaced to cause a tracking radar with closed-loop tracking to align its antenna axis at a point many times the scatterer spacing away from the scatterers. If the scatterers are stationary, the radar antenna will stay pointing in the erroneous direction. Figure 9.20 shows experimental data demonstrating this phenomenon with a two-reflector target.

The angle noise phenomenon affects all types of tracking radars but is mainly of concern for tracking radars where precision target location is needed. To aid in visualizing why angle noise affects any radar-type angular-direction-sensing device, the echo signal propagating in space was analyzed, showing that the angle noise is present in this propagating energy as a distortion of the phase front. Theoretical plots of a distorted phase front from dual sources compare very closely with photographs of the phase front of the radiating surface ripples in the ripple-tank experiment with dual vibrating probes.[37] All radar angle-sensing devices sense, by one means or another, the phase front of the signal and indicate the target to be in a direction normal to the phase front. Thus, the phase-front distortions affect all types of angle-sensing radars.

Many measurements of angle noise have been made on a variety of aircraft, and the results of theoretical studies have been verified. The theory and measurements show that angle noise expressed in linear units of displacement, such as meters, of the apparent position of the target from the center of gravity of the target is independent of range (except for very short ranges). Therefore, rms angle noise σ_{ang} is expressed in

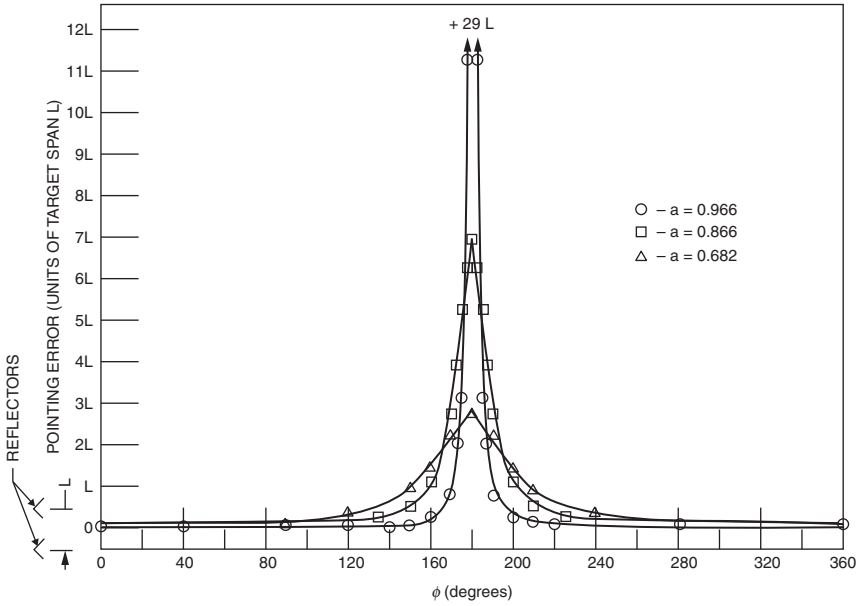


FIGURE 9.20 Apparent location of a dual-source target as a function of relative phase ϕ for different values of relative amplitude a measured with a tracking radar (*Figure 5 from Howard*[37])

units of meters of error measured at the target location. The results show that the rms value of angle noise σ_{ang} is equal to $R_o/\sqrt{2}$, where R_o is the radius-of-gyration* (taken along the angular coordinate of interest) of the distribution of the reflecting areas of the target.[33] For example, if a target's reflecting areas have a $\cos^2(\pi\alpha/L)$-shaped distribution, where α is a variable and the target span is from $+L/2$ to $-L/2$, calculation of the radius of gyration divided by $\sqrt{2}$ gives a value of σ_{ang} of $0.19L$. Typical values of σ_{ang} on actual aircraft fall between $0.15L$ and $2.0L$, depending upon the distribution of the major reflecting areas such as engines, wing tanks, and so on. A small aircraft, nose-on view, with a single engine and no significant reflectors attached to the wings will have a σ_{ang} of approximately $0.1L$, whereas larger aircraft with an outboard engine and possibly wing tanks will have a σ_{ang} approaching the value of $0.2L$. The aircraft side view also tends toward the value of $0.2L$ because of a more continuous distribution of reflecting areas. Estimation of angle scintillation rms error in units of target span can be made by relating the approximate target distribution in Figure 9.21 with actual aircraft configurations.

The value of σ_{ang} for a complex target is essentially a fixed value regardless of RF frequency, if a target span of at least several wavelengths is assumed and is independent of the rate of random motion of the target. However, as described later, the spectral distribution of angle-noise power is directly affected by radar frequency, atmospheric turbulence, and other parameters.

* Radius-of-gyration is calculated assuming the "weight" of the scatterers is their effective radar scattering cross section.

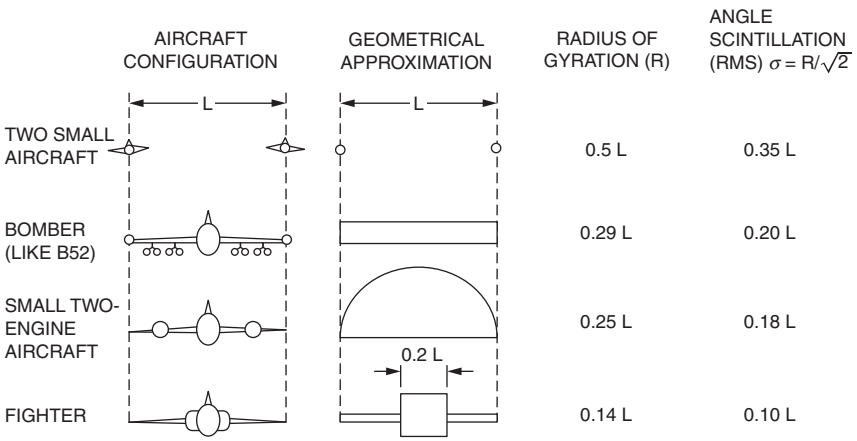


FIGURE 9.21 RMS angle scintillation based on the theoretical relation to the radius-of-gyration of the distribution of reflecting areas of the target

Target angle noise is typically gaussian-distributed. An example of the measured distribution of the apparent target angle of a small two-engine aircraft is shown in Figure 9.22. A relatively long time sample is needed, since short time samples of data can depart from the gaussian shape. Unusual targets may also depart from gaussian distributed angle noise. Delano[38] gives data from two aircraft in formation that are gaussian-distributed when completely unresolved, but change shape at close range where the antenna begins to resolve the two aircraft (as described in Section 9.11).

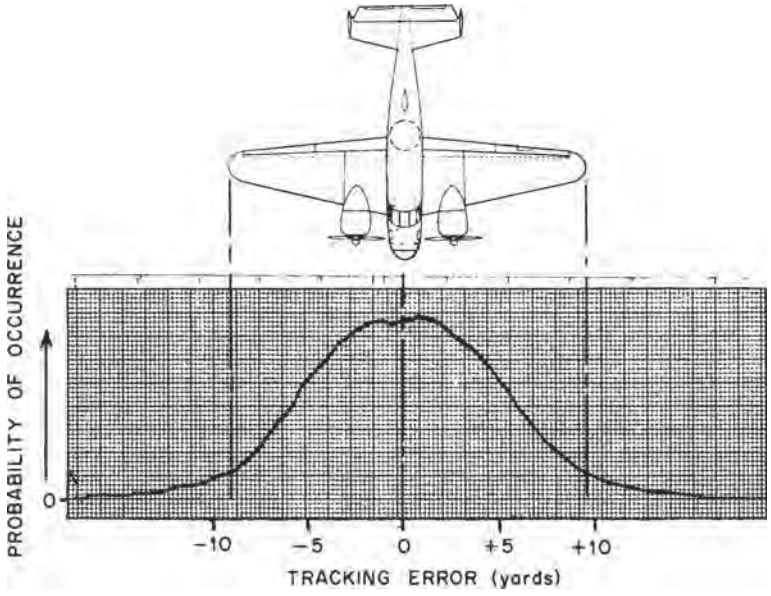


FIGURE 9.22 Amplitude probability distribution of angle scintillation measured on a small two-engine aircraft

Although the rms value of angle noise is essentially a constant for a given target and aspect, the spectral distribution of this energy is dependent on radar frequency and the random target motion. A typical spectrum shape is

$$N(f) = \sigma_{\text{ang}}^2 \frac{2B}{\pi(B^2 + f^2)} \tag{9.6}$$

where $N(f)$ = spectral noise power density, power/Hz
B = noise bandwidth, Hz
f = frequency, Hz

The values of B are proportional to radar frequency and dependent upon air turbulence effects on target motion and target aspect. An example of a measured angle scintillation spectrum is shown in Figure 9.23. Typical values of B at X band, in relatively turbulent air, range from about 1.0 Hz for small aircraft to about 2.5 Hz for larger aircraft. B changes in proportion to radar frequency provided that the target span is at least a few wavelengths. Again, long time samples are necessary to obtain a relatively smooth spectrum from measured data. For the above values of B, about 7 minutes of data was necessary to reach essentially the long-time-averaged characteristic. This is a reference point about which there will be considerable variation for a typical time period of interest. For example, with only 1 minute of data the noise power σ_{ang} would vary over 0.5 to 1.5 times the long-time-averaged σ_{ang}. At lower radar frequencies and in less turbulent atmosphere, B may be smaller, and proportionately longer time samples are necessary; thus, for short time samples of radar performance, significant statistical variations must be expected.

To convert σ_{ang} expressed in linear units measured at the target to angular units for a radar at range r, the following relation may be used:

$$\sigma_{\text{ang}}\,(\text{angular mils}) = \sigma_{\text{ang}}\,(\text{m})/r(\text{km})$$

Because the angular errors caused by angle noise are inversely proportional to range, angle noise is of concern mainly at medium and close ranges. The resultant tracking noise can be reduced by lowering the servo bandwidth to reduce the radar's ability to follow the higher-frequency components of the noise. The amount of noise reduction may be estimated by comparing the area under a spectral-power-density plot of angle noise below the frequency corresponding to the radar servo bandwidth with the total area under the power-density plot. (The spectral-power-density plot may be obtained by squaring the ordinate values of a spectral-distribution plot such as shown in Figure 9.23).

The choice of AGC characteristics also affects the amount of angle noise followed by a tracking antenna. The AGC voltage is generated from the sum signal and follows the echo-signal-amplitude fluctuation. There is a degree of correlation between the angle-noise magnitude and echo-signal magnitude such that angle-noise peaks are generally accompanied by a dip or fade in amplitude. A slow AGC system that does not maintain constant signal level during rapid changes allows the signal level to drop during a rapid fade, reducing sensitivity (volts per degree angle error) during the large angle-noise peaks. This results in a smaller rms tracking noise with a slow AGC system.[39,40]

However, this reasoning neglects an additional noise term, caused by the lack of full AGC action, which is proportional to tracking lag. A tracking lag causes a dc error

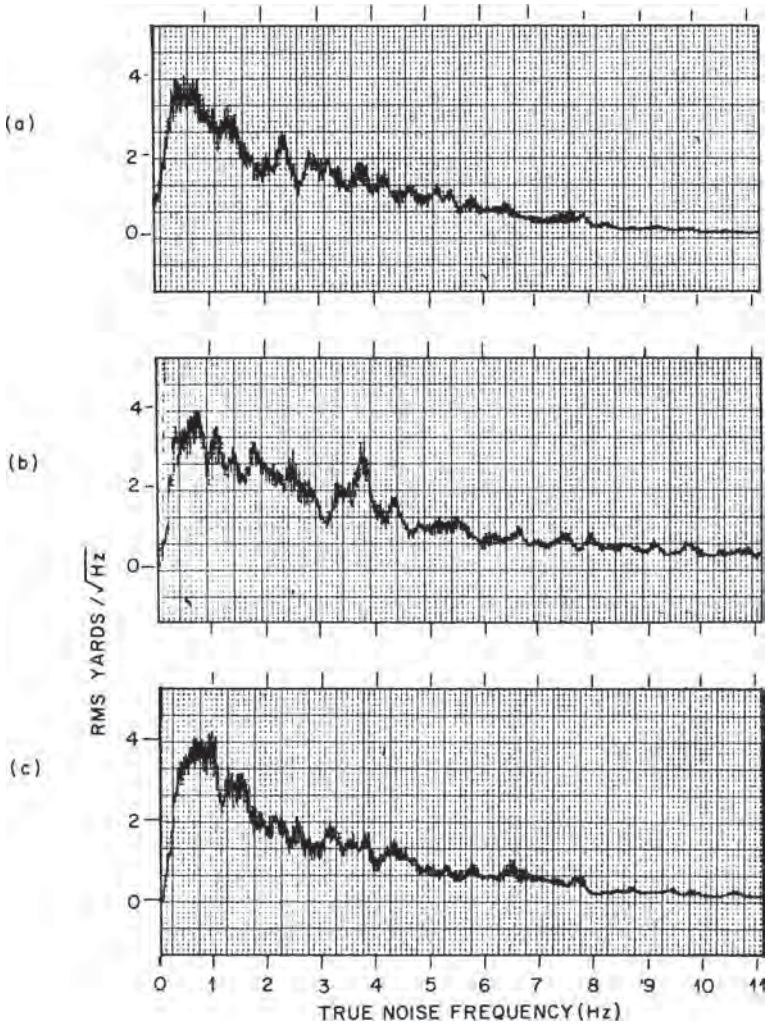


FIGURE 9.23 Spectral-energy distribution of angle scintillation measured on the nose aspect of a small two-engine aircraft

voltage in the angle-error-detector output equal to angle error times the angle sensitivity. A slow AGC allows the amplitude noise to modulate the true tracking-error voltage, causing additional noise in angle tracking. Thus, there will be an additional rms angle error proportional to tracking lag and dependent on the AGC time constant,[40] as illustrated in Figure 9.24.

In general, a fast AGC is recommended because it reduces the additional noise term allowed by slow AGC and the possibility of larger rms tracking errors, which can be considerably greater than the angle noise with a fast AGC. As previously discussed, angle noise is significant, mainly at medium and close range where target angle rates are greatest. As seen in Figure 9.24 a tracking lag of only one-half the

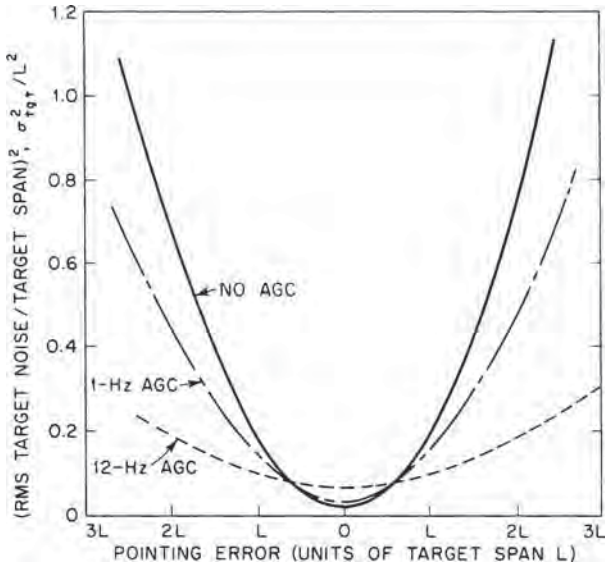


FIGURE 9.24 Angle-scintillation noise power as a function of tracking error for three different AGC bandwidths (*from Dunn, Howard, and King*[33] *© IRE 1959*)

target span will result in greater tracking noise in a slow AGC system, with the danger of much higher noise with greater lag. Therefore, for overall performance, a fast AGC is recommended.

Range Noise (Range Glint). Range noise, or random tracking errors in the range coordinate caused by complex targets, is a significant basic limitation in range tracking. Acquisition of a desired spectral line by a doppler frequency tracking system is also limited by range noise. Coarse velocity information is obtained by differentiation of range to determine the desired spectral line. Range noise is a major limitation to the accuracy of velocity obtained from range rate and can prevent selection of the desired spectral line.

The range-tracking errors caused by a finite-size target and by multipath also cause significant angle-tracking errors in multilateration tracking systems that triangulate using high-precision range measurements from multiple locations to calculate target angle location. Multilateration systems, such as the Pacific Missile Range Extended Area Tracking System (EATS), depend upon very precise range measurements. Small range-tracking errors cause significant errors in calculated target angle based on the range measurements. These errors must be fully understood to assess the performance of multilateration systems.

Target-caused range-tracking errors, similar to target-caused angle errors, are greater than the wander of the target center of gravity and can fall outside the target span.[41] Figure 9.25 shows typical samples of spectral-energy distributions and probability density functions for different target configurations. The range noise measurements were made on small and large aircraft and multiple aircraft using the split video range error detector.[42] The characteristics follow very closely to the relations of target angle noise to the target configuration radius-of-gyration along the angle coordinate.

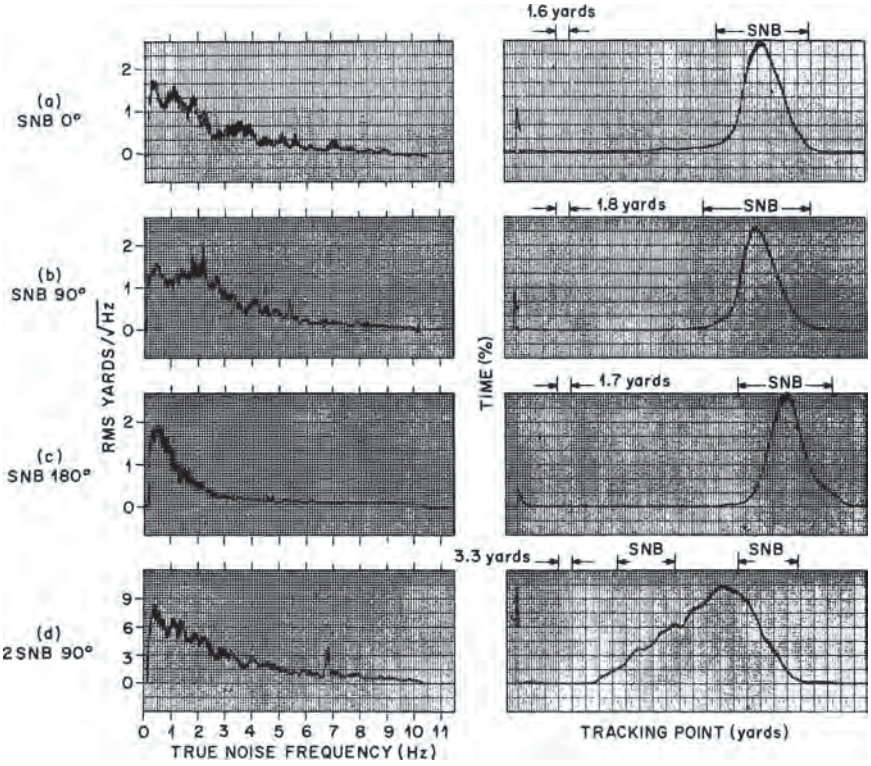


FIGURE 9.25 Typical spectral-energy distributions for the three views of the SNB aircraft: (*a*) nose view, (*b*) side view, (*c*) tail view, and (*d*) side view of two small two-engine aircraft flying in formation

For range tracking, it is necessary to relate range noise to the target reflectivity distribution along the range coordinate. In general, the long-time-average value of the rms range error may be closely estimated by taking 0.8 times the radius-of-gyration of the distribution of the reflecting areas in the range dimension based on many measurements of small, large, and multiple aircraft. Typically, in terms of target span along the range coordinate, the rms value will fall between 0.1 and 0.3 times the target span—being close to the multiplier, 0.3, for the tail view and nose view, and 0.1, for the side view. The spectral shape may be closely estimated by using the same function of frequency as described for angle noise and the same value of bandwidth. The error as a function of relative phase and amplitude of the target reflectors is similar to the angle noise phenomenon.

A beacon on a cooperative target can provide a point source (single-pulse response) to eliminate range error caused by the target. However, very stable circuitry is required to avoid pulse jitter and drift.

Doppler Scintillation and Spectral Lines. Doppler scintillation and spectral lines caused by a complex target may be divided into two phenomena[43]: (1) spectral lines caused by parts of the aircraft such as propellers and jet turbine blades, and (2) a continuous doppler spectrum spread by the motion of an aircraft in flight symmetrically

above and below the average doppler of the target. A target typically has a significant random yaw, pitch, and roll motion even on a "fixed" heading. Time plots of typical aircraft heading for an aircraft flying a "straight course" are observed to have typical random yaw motion that causes small changes in the doppler from each of the scattering surfaces of the aircraft's rigid structure. Relative to the average doppler of the aircraft, the scattering surfaces located away from the aircraft center will have a small increasing and decreasing relative doppler frequency as the aircraft yaws right and left. This causes a spectral spread of the doppler of the echo from the rigid body of the aircraft and is accompanied by spectral lines caused by moving parts on the aircraft.

Components of the target echo from rotating or moving parts of the target cause doppler lines at frequencies displaced from the airframe doppler spectrum. The periodic amplitude modulation causes pairs of doppler lines symmetrical about the doppler of the airframe velocity. Moving parts can also cause pure frequency modulation that will result in a single set of doppler lines on one side of the airframe doppler spectrum.[43]

A major significance of the doppler modulation is its effect on doppler-measuring radars. A doppler tracking system that automatically tracks the frequency of a spectral line of the echo is subjected to two problems: (1) there is the possibility of locking on a false line caused by moving parts of the target; and (2) when properly locked onto the airframe doppler spectrum, the doppler reading will be noisy as defined by the random fluctuation in instantaneous frequency as observed by the spread of the doppler spectrum. Coherent beacons (which receive, amplify, and transmit received radar pulses) can provide a doppler-shifted response free of target-caused spectral spread and periodic modulations. A delay time is provided to separate the beacon response from the target echo.

Target doppler scintillation also offers useful information about the target configuration. Normal target motion will result in different doppler shifts for each major scatterer of a rigid-body target, and the shifts will be a function of the displacement of the scatterer from a reference point such as the center of rotation of the target's random motions. Therefore, a high-resolution doppler system can resolve major reflectors and locate them in cross range as a function of the doppler difference from the reference reflector. This technique, called *inverse synthetic aperture radar (ISAR),* uses the target motion for the needed aspect change, instead of radar motion as used in conventional synthetic aperture radar, to obtain detailed cross-range target image information.[44,45]

9.9 OTHER EXTERNAL CAUSES OF ERROR

Multipath. Multipath angle errors result from reflections of the target echo from objects or surfaces causing echo pulses to arrive by other than the direct path to the radar beam in addition to the direct path. These errors are sometimes called *low-angle tracking errors* when applied to tracking of targets at small elevation angles over the Earth or ocean surface.[46–50] Multipath errors are typically a special dual-source case of angle noise resulting from the geometry as described in Figure 9.26, where the target and its image reflected from a

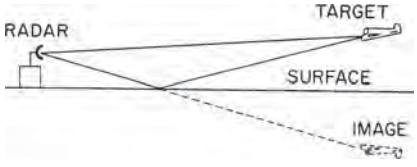


FIGURE 9.26 Geometry of the radar multipath tracking condition, where the reflection from a surface appears to the radar as an image below the surface

surface to the radar beam are the two sources. Over a smooth ocean surface, they are separated only in the elevation coordinate so that most of the error appears in the elevation-tracking channel. Severe elevation errors may cause some cross coupling of the error to the azimuth channel. Rough surfaces cause diffuse scattering, which can contribute errors to both azimuth and elevation tracking.[46] Different path geometries such as non-flat land or a building may also cause a significant error to appear in the azimuth-tracking channel.

The major difficulty with low-angle tracking is that the target and its image are essentially coherent and their relative phase changes slowly and the angular error it causes is readily followed by an angle-tracking system. Furthermore, the paths are almost equal, and in most cases, they cannot be resolved by high-range-resolution techniques. Long time averages of the data do not, in practice, give target elevation; thus, the multipath angle error has no simple solution and is generally minimized by using narrow-beam antennas.

When the target is at low altitude, the multipath errors are severe, as observed in the measured data shown in Figure 9.27. This data is the multipath error of a 2.7° beamwidth S band (3-GHz) tracking radar that is tracking an aircraft target with a beacon at 3300-ft altitude. An AN/FPS-16 tracking radar with a 1.1° beamwidth at C band (5.7 GHz) was used to simultaneously track with its narrow beam, which remained above the sea surface without significant multipath error, to provide a true target altitude reference for the data in Figure 9.27. There is a measurement bias error (observed in Figure 9.27) of about 0.25°.

The tracking data from a radar tracking a target with a beacon is plotted in Figure 9.27, showing a typical multipath error illustrating the phenomenon from the region where the image enters the sidelobes to the region where it enters the main beam. There are three methods used for predicting multipath errors, depending upon where the reflected target image enters the antenna pattern. At the far range,

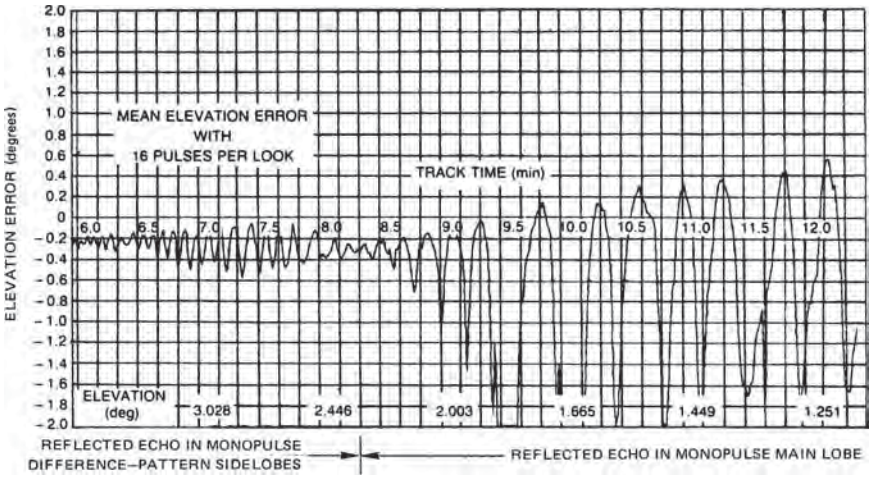


FIGURE 9.27 Measured elevation-tracking error of an S-band radar using an AN/FPS-16 radar for a target elevation reference

the image enters the antenna main beam, and the error is essentially that of a two-reflector target glint error following approximately the equation[37]

$$e = 2h\frac{\rho^2 + \rho\ \cos\phi}{1 + \rho^2 + 2\rho\ \cos\phi} \tag{9.7}$$

where e = error, same units as h, measured at the target range relative to the target
ρ = magnitude of surface reflection coefficient
h = height of target
ϕ = relative phase determined by geometry of direct and surface-reflected signal paths, as shown in Figure 9.26.

Although the fluctuations in ρ and ϕ alter the actual tracking from the theoretical, the equation gives a good indication of the errors to be expected when tracking a point source such as a beacon. However, skin tracking of an aircraft at low elevation may result in a departure from the classic periodic error versus elevation, as illustrated in Figure 9.27, because of an interaction between target angle scintillation and multipath error that can change the characteristics of the multipath error.

When tracking a point source target at close range, the radar main beam is above the image, but the image is seen by the difference-pattern sidelobes. The multipath errors that result are cyclic, almost sinusoidal, with an rms value predicted by the equation[46]

$$\sigma_E = \frac{\rho\theta_B}{\sqrt{8G_{se}(\text{peak})}} \tag{9.8}$$

where σ_E = rms elevation angle multipath error, same units as θ_B
θ_B = one-way antenna beamwidth
ρ = reflection coefficient

and G_{se} (peak) is the power ratio of the tracking-antenna sum-pattern peak to the error-pattern peak sidelobe level at the angle of arrival of the image signal.

The cyclic rate may be approximated by the equation

$$f_m = \frac{2hE}{\lambda} \tag{9.9}$$

where f_m = frequency of cyclic multipath error, rad/s
h = height of radar antenna
λ = wavelength, same units as h
E = rate of target elevation change as seen by radar, rad/s

The intermediate range is between the short-range region where the image appears in the sidelobes, and the long-range region, where the image appears within the half-power beamwidth. The error is difficult to calculate in this region because it falls in the nonlinear error-sensing portion of the antenna patterns, and the radar response is strongly dependent upon the specific feed design and error-processing technique. However, Figure 9.28[46,47] provides a practical means for approximating multipath-error values in this region. The curves are calculated multipath errors based on an assumed gaussian-shaped sum pattern and derivative of the sum pattern

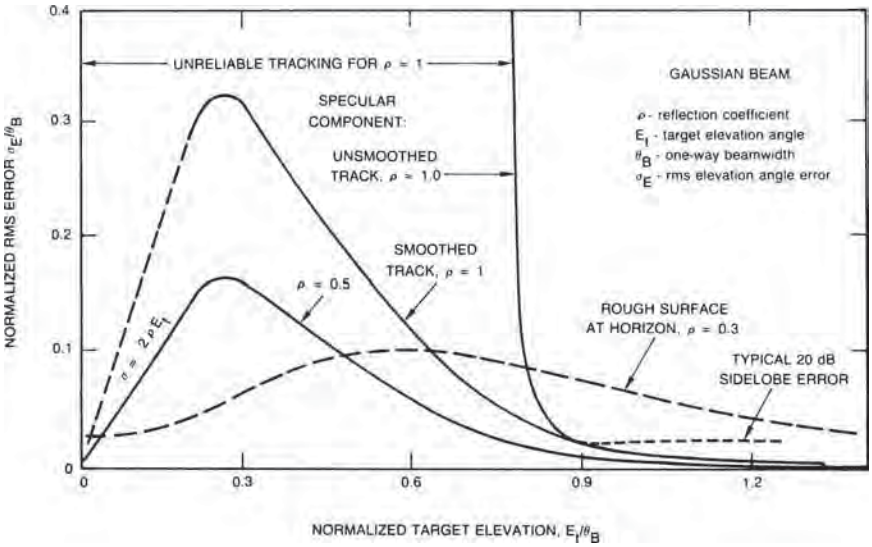


FIGURE 9.28 Calculated rms multipath error σ_E versus target elevation E_r, both normalized to radar beamwidth θ_B

as the monopulse difference pattern. Figure 9.28 shows typical sidelobe multipath errors for higher-elevation targets and the linearly decreasing error versus target elevation, predicted by the above equation, for very low-elevation targets. The graph is normalized to radar beamwidth on both axes for convenient use with a wide variety of radars. The dashed portions of the curves are regions of uncertainty because of significant variations of reflection for a given sea state.

In the intermediate region, the error increases to a peak at target elevations of about 0.3 beamwidth. The peak value is dependent on several factors including surface roughness (which, in part, determines the value of ρ), servo bandwidth, and antenna characteristics in the region. The errors are severe, and with *un-smoothed track* (wide servo bandwidth), the radar can break lock and lose track of the target.

When the surface is rough, corresponding to a reflection coefficient of about 0.3, the characteristic of the error versus elevation changes is observed in Figure 9.28. The rough surface causes significant diffuse scattering rather than a mirror reflection. This changes the shape of the error curve and results in some residual elevation-angle error when the target elevation goes to zero. It also causes some significant azimuth error.[46]

Crosstalk Caused by Cross-Polarized Energy. Target echo energy cross-polarized to the radar antenna causes crosstalk (cross coupling) in radars; i.e., the azimuth error causes output from the elevation-error detector, and the elevation error causes output from the azimuth-error detector. Generally, this effect is negligible because the cross-polarized energy is usually a small fraction of the received polarization from typical targets, and it is normally reduced by about 20 dB by the antenna design. However, in special cases, the resultant crosstalk can be very high and may cause a large tracking error and possible loss of track. For example, polarization of a linearly polarized beacon on a target could rotate with target aspect change and, in the worst case, approach a cross-polarized condition.

Theoretically, the coupling to cross-polarized energy is zero when the source is precisely on axis and increases with displacement from the axis.[51] The cross-polarization error in a tracking radar system is pure crosstalk so that a small true tracking error in one tracking coordinate causes the antenna to move in the other coordinate. The error in the second coordinate then causes the antenna to move farther from the source in the first coordinate. When there is no retarding effect, the cross-polarized energy causes the antenna to drive off target in one of the quadrants of the two-axis angle-tracking coordinate system, depending upon the direction of the initial error that moved the source off the precise on-axis position.[4,51]

A solution used with missile-range-instrumentation radar, where target aspect changes can cause a linearly polarized beacon to rotate to a cross-polarized aspect, is to provide a circular polarization tracking capability. Coupling a linearly polarized signal to a circularly polarized antenna results in a 3 dB signal loss, but it is independent of the direction of the linear polarization when rotated about a line in the direction toward the radar.

Troposphere Propagation. The troposphere is typically a nonhomogeneous medium for propagation and will cause random beam bending. Figure 9.29 illustrates the approximate relation of rms angle error to various atmospheric conditions.[22] The worst case is heavy cumulus clouds, which cause columns of air, shaded from the sun by the clouds, that are cooler than the surrounding air and consequently of a different dielectric constant. The result is typically a random beam bending as the radiated energy passes through these columns. Figure 9.29 applies only for the portion

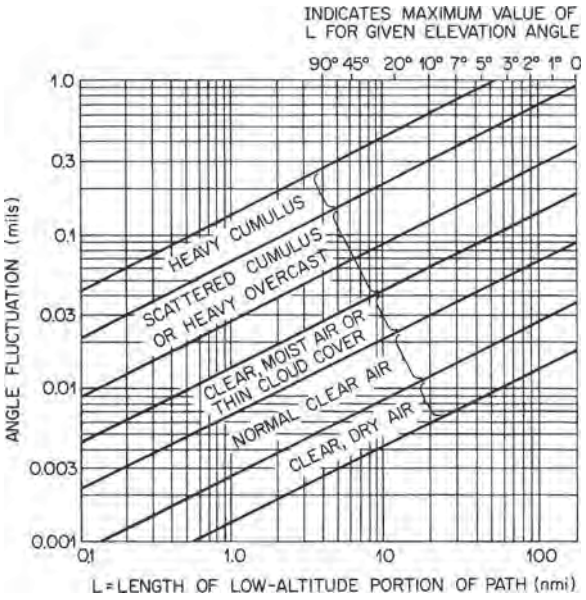


FIGURE 9.29 Angle fluctuation versus path length for different tropospheres (*from Final Report: Instrumentation Radar AN/FPS-16 (XN-2) by RCA under contract Bu Aer NOas 55-869c*)

of the beam that is within the troposphere. Once the beam goes above the troposphere (typically about 6 to 9 km), there is no further beam bending.

The troposphere also affects target range measurement, but the errors are small, in the order of 0.3 to 0.6 m maximum. However, even small errors of this magnitude will cause significant errors in multilateration systems that determine target angle by calculations using precise range measurements from separate locations.

9.10 INTERNAL SOURCES OF ERROR

Receiver Thermal Noise. The angle error caused by receiver thermal noise in a monopulse tracking system is

$$\sigma_t = \frac{\theta_B}{k_m \sqrt{B\tau(S/N)(f_r/\beta_n)}} \tag{9.10}$$

where k_m is the angle-error-detector slope. The value of k_m is determined by the steepness of the antenna difference patterns, and a variety of values can be obtained, depending on the type of feed used. The values vary from 1.2 for the original four-horn square feed to a maximum of 1.9 for the MIT 12-horn feed. However, as described in the discussion on feeds, the 12-horn feed gives a lower antenna efficiency (0.58) than an optimum multimode monopulse feed, which can approach an efficiency of 0.75, although its angle sensitivity is less, typically having a value of 1.7. Therefore, there is a tradeoff between slope and efficiency. A slope for monopulse radars is dependent upon feed design and is typically 1.57 for a good modern four-horn multimode feed design.

When there is a significant tracking lag or deliberate beam offset from the target, the error σ_{to}, due to receiver noise for a given SNR, is given by the equation

$$\sigma_{t0} = \sigma_t \sqrt{L[1 + k(\theta_L/\theta_B)^2]} \tag{9.11}$$

where θ_L = lag angle, same units as θ_B
L = antenna sum-pattern loss at angle θ_L

A similar range-tracking error σ_{rt} results from receiver noise. The equation relating the error to the SNR and system parameters is

$$\sigma_{rt} = \frac{\tau}{\sqrt{k_r (S/N) f_r/\beta_n}} \quad \text{ft (rms)} \tag{9.12}$$

where τ = pulse length in ft
k_r = range-error-detector sensitivity (maximum value of 2.5 for a receiver where $B = 1.4$)
S/N = signal-to-noise ratio
β_n = servo bandwidth

Other Internal Sources of Error. There are many other sources of internal errors that are small in well-designed tracking radars. These include changes in relative phase and amplitude between monopulse receiver channels as a function of signal strength,

TABLE 9.1 Inventory of Range-Error Components*

Component	Bias	Noise
Radar-dependent tracking errors	Zero range setting Range discriminator shift servo unbalance Receiver delay	Receiver thermal noise Multipath Servo electrical noise Servo mechanical noise Variation in receiver delay
Radar-dependent translation errors	Range oscillator frequency Data takeoff zero setting	Range resolver error Internal jitter Data gear nonlinearity and backlash Data takeoff nonlinearity and granularity Range oscillator instability
Target-dependent tracking errors	Dynamic lag Beacon delay	Dynamic lag Glint Scintillation Beacon jitter
Propagation error	Average tropospheric refraction Average ionospheric refraction	Irregularities in tropospheric refraction Irregularities in ionospheric refraction

*From D. K. Barton in "Modern Radar," R. S. Berkowitz (ed.), New York: John Wiley & Sons, 1965, chap. 7, p. 622.

RF frequency, detuning, and temperature. Also, pedestal bending from solar heating, nonorthogonality of pedestal axes, gearing backlash, bearing wobble, granularity of data readout, and many other factors contribute to errors. Table 9.1 lists the magnitude of these errors for the precision instrumentation radar AN/FPS-16.[22]

Calibration is important to minimize errors.[22] When maximum performance is required, timely accurate calibration must be performed. The procedure may require up to four hours to fully stabilize the radar system. For instrumentation radar, where the time of a tracking event is known, final calibration is performed just preceding the event to minimize drift errors.

9.11 SUMMARY OF SOURCES OF ERROR

Angle Measurement Errors. An inventory of angle measurement errors is given in Table 9.2. This includes several sources of error that should be considered in addition to the radar-related sources.

Figure 9.30 is an example of the measured tracking performance of an AN/FPS-16 radar tracking a 6-in metal sphere that provides a point source target to eliminate target-caused errors. The data illustrates which error sources dominate at different regions of the radar range and their characteristics versus range.

The target-caused errors discussed in Section 9.8 include the usual tracking events where the target extent is within the 3-dB beamwidth of the radar. However, a large target such as an aircraft formation may extend beyond the linear angle-error region of the antenna patterns and eventually reach the point of resolution of one of the aircraft. The resultant angle-tracking error for large targets is illustrated by the example in Figure 9.31. In Figure 9.31*a,* the typical gaussian-like glint error distribution is observed. With the wider separation of the aircraft, the tracking-error

TABLE 9.2 Inventory of Angle-Error Components*

Component	Bias	Noise
Radar-dependent tracking errors (deviation of antenna from target)	Boresight axis collimation Axis shift with RF and IF tuning Receiver phase shift Target amplitude Temperature Wind force Antenna unbalance Servo unbalance	Receiver thermal noise Multipath (elevation only) Wind gusts Servo electrical noise Servo mechanical noise
Radar-dependent translation errors (errors in converting antenna position to angular coordinates)	Leveling of pedestal North alignment Static flexure of pedestal and antenna Orthogonality of axes solar heating	Dynamic deflection of pedestal and antenna Bearing wobble Data gear nonlinearity and backlash Data takeoff nonlinearity and granularity
Target-dependent tracking errors	Dynamic lag	Glint Dynamic lag variation Scintillation Beacon modulation
Propagation errors	Average refraction of troposphere Average refraction of ionosphere	Irregularities in tropospheric refraction Irregularities in ionospheric refraction
Apparent or instrumentation errors (for optical reference)	Telescope or reference instrument stability Film emulsion and base stability Optical parallax	Telescope, camera, or reference instrument vibration Film-transport jitter Reading error Granularity error Variation in optical parallax

*From D. K. Barton in "Modern Radar," R. S. Berkowitz (ed.), New York: John Wiley & Sons, 1965, chap. 7, p. 618.

distribution changes shape, becoming somewhat rectangular with a separation of aircraft as in Figure 9.31*b*. At the widest separation, where the aircraft are almost resolved, as in Figure 9.31*c*, the radar will track one aircraft until it fades and the other aircraft blossoms in amplitude. Then the radar-tracking point will move to the other aircraft. The dwell on each target with random switching between the two aircraft causes the double-humped distribution of error.

Range Measurement Errors. The major sources of target range-error measurement errors are given in Table 9.3. Typical bias and noise of target range measurement errors in a precision-ranging radar are equal to a total rms value of 1.6 m rms. Further details of range-error sources and their magnitude are give in Section 10.3 of Barton.[22]

Limitations of Performance. Mitchell et al.[52] describe basic performance limitations of the AN/FPQ-6 high precision tracking radar measured under ideal

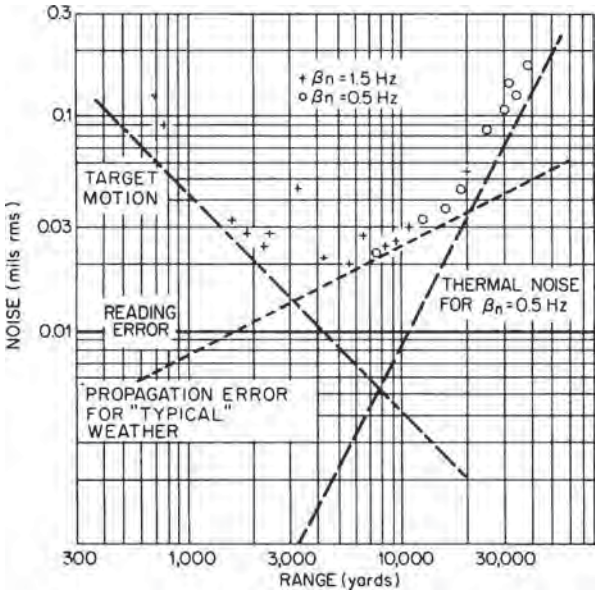


FIGURE 9.30 Azimuth-tracking noise versus range using a 6-in metal sphere target supported inside a balloon to minimize target motion (estimated at 1.5 in rms). β_n = servo bandwidth

conditions with a carefully designed boresighting facility. This task provided measured data verifying the anticipated performance of the highest precision tracking radar at that time.

TABLE 9.3 Inventory of Range-Error Components*

Component	Bias	Noise
Radar-dependent tracking errors	Zero range setting Range discriminator shift servo unbalance Receiver delay	Receiver thermal noise Multipath Servo electrical noise Servo mechanical noise Variation in receiver delay
Radar-dependent translation errors	Range oscillator frequency Data takeoff zero setting	Range resolver error Internal jitter Data gear nonlinearity and backlash Data takeoff nonlinearity and granularity Range oscillator instability
Target-dependent tracking errors	Dynamic lag Beacon delay	Dynamic lag Glint Scintillation Beacon jitter
Propagation error	Average tropospheric refraction Average ionospheric refraction	Irregularities in tropospheric refraction Irregularities in ionospheric refraction

*From D. K. Barton in "Modern Radar," R. S. Berkowitz (ed.) New York: John Wiley & Sons, 1965, chap. 7, p. 622.

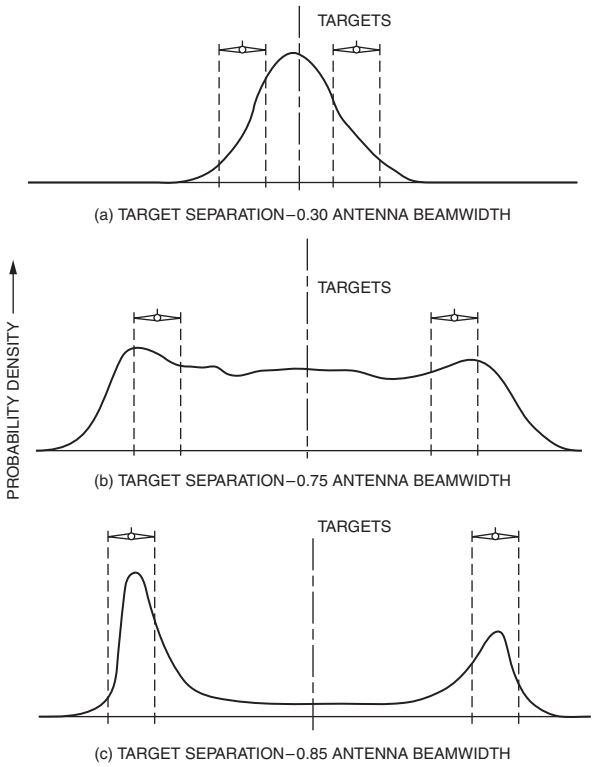


FIGURE 9.31 Probability distribution of radar pointing when tracking two targets (where the left target is approximately 1.5 dB larger than the right target). Three different angular separations of the targets are (*a*) 0.30 antenna beamwidth, (*b*) 0.75 antenna beamwidth, and (*c*) 0.85 antenna beamwidth.

9.12 ERROR REDUCTION TECHNIQUES

Multipath-Error Reduction. Very-low-altitude targets cause severe elevation-angle tracking errors, as described in Section 9.9, which may result in useless elevation tracking data and possible loss of tracking of the target. A variety of techniques have been developed to reduce these errors or their effects on radar tracking.[53–58] One simple approach to avoid loss of tracking in elevation is to open the elevation-tracking servo loop and place the antenna beam at about a half beamwidth above the horizon.[1] Azimuth closed-loop tracking may continue. Although the elevation-angle-error-detector output has large indicated angle errors, it is monitored to observe whether or not the target is maneuvering upward through the beam. A target rising through the beam will cause a positive angle-tracking error indication and the closed-loop elevation tracking resumes.

A very effective and direct approach to multipath-error reduction is to use a very narrow beam, usually accomplished by operating at short wavelengths such as an 8-mm (35 GHz or K_a band) region with the usual microwave-tracking aperture size.[13,25,26,

This approach can reduce errors by two effects. First, as observed in Figure 9.28, the magnitude of the elevation multipath error reduces in direct proportion to the beamwidth. The second advantage of shorter wavelengths is that even a relatively smooth sea, such as sea state 1, has wave heights of many wavelengths and appears rough, resulting in a smaller reflection coefficient. This is observed in Figure 9.28 to give small multipath errors. The 8-mm-wavelength monopulse capability may be effectively combined with a lower microwave band as described in Section 9.6 to take advantage of the complementary features of both bands.

Target Angle and Range Scintillation (Glint) Reduction. Target-caused errors in angle and range tracking may be reduced by filtering, such as reducing tracking servo bandwidth. However, sufficient servo bandwidth must be retained to follow target trajectories. Unfortunately, target angle and range scintillation power density is normally concentrated below about 1 to 2 Hz when operating at microwave bands and falls within normally required bandwidths.

Target scintillation total noise power is relatively independent of frequency, but the spectral energy tends to spread upward in frequency as wavelength is reduced, resulting in lower noise power density in the servo passband. Therefore, operating at a shorter wavelength will result in lower target noise effects on closed-loop tracking.

Diversity techniques, which can provide statistically independent samples of target scintillation, offer a means for reducing target scintillation effects. The most practical technique is frequency diversity using pulse-to-pulse radar frequency change, which will alter the phase relations between the echoes from dominant reflecting surfaces of the target.[59–62] The frequency change must be sufficient to cause enough change in relative phases of the reflectors to result in statistically independent samples of target scintillation at each new frequency. An approximate rule is a minimum frequency change of $1/\tau$, where τ is the radar range delay time between the leading and lagging extremities of the target. The reduction in rms angle and range target scintillation may be approximated by dividing by the square root of n, where n is the number of frequency steps provided.

Reduction of Internally Caused Errors. Angle errors caused by receiver thermal noise, as well as target scintillation, are minimized by maintaining the target as closely as possible to the tracking axes. The technique called *on-axis tracking,* described in Section 9.6, is a means of placing a computer in the tracking loop to minimize lag and provide optimum angle-error filtering.

Accurate system calibration also greatly reduces internal sources of error. Frequent calibration corrects for drift in component gain and phase and pedestal flexure. Other internal sources of error with known characteristics can be automatically corrected to minimize their contamination of the output data.

REFERENCES

1. S. M. Sherman, *Monopulse Principles and Techniques*, Norwood, MA: Artech House, 1986.
2. A. I. Leonov and K. I. Formichev, *Monopulse Radar*," Norwood, MA: Artech House, 1986.
3. J. H. Dunn and D. D. Howard, "Precision tracking with monopulse radar," *Electronics*, vol. 33, pp. 51–56, April 22, 1960.
4. P. Z. Peebles, Jr., "Signal Processor and accuracy of three-beam monopulse tracking radar," *IEEE Trans*., vol. AES-5, pp. 52–57, January 1969.

5. P. W. Hannan, "Optimum feeds for all three modes of a monopulse antenna, I: Theory; II: Practice," *IEEE Trans*., vol. AP-9, pp. 444–460, September 1961.
6. "Final report on instrumentation radar AN/FPS-16 (XN-2)," Radio Corporation of America, unpublished report NTIS 250500, pp. 4-123–4-125.
7. D. K. Barton, "Recent developments in radar instrumentation," *Astron. Aerosp. Eng*., vol. 1, pp. 54–59, July 1963.
8. J. T. Nessmith, "Range instrumentation radars," *IEEE Trans*., vol. AES-12, pp. 756–766, November 1976.
9. J. A. DiCurcio, "AN/TPQ-27 precision tracking radar," in *IEEE Int. Radar Conf. Rec*., Arlington, VA, 1980, pp. 20–25.
10. D. D. Howard, "Single Aperture monopulse radar multi-mode antenna feed and homing device," in Proc. *IEEE Int. Conv. Mil. Electron. Conf*., September 14–16, 1964, pp. 259–263.
11. P. Mikulich, R. Dolusic, C. Profera, and L. Yorkins, "High gain cassegrain monopulse antenna," in *IEEE G-AP Int. Antenna Propag. Symp. Rec*., September 1968.
12. R. C. Johnson and H. Jasik, *Antenna Engineering Handbook*, 2nd Ed., New York: McGraw-Hill Book Company, 1984, Chap. 34.
13. D. Cross, D. Howard, M. Lipka, A. Mays, and E. Ornstein, "TRAKX: A dual-frequency tracking radar," *Microwave J*., vol. 19, pp. 39–41, September 1976.
14. V. W. Hammond and K. H. Wedge, "The application of phased-array instrumentation radar in test and evaluation support," in *Electron. Nat. Security Conf. Rec*., Singapore, January 17–19, 1985.
15. J. W. Bornholdt, "Instrumentation radars: Technical evaluation and use," in *Proc. Int. Telemetry Council*, November 1987.
16. W. B. Milway, "Multiple target instrumentation radars for military test and evaluation," in *Proc. Int. Telemetry Conf*. vol. XXI, 1985.
17. R. L. Stegall, "Multiple object tracking radar: System engineering considerations," in *Proc. Int*. Telemetry Council, 1987.
18. R. S. Noblit, "Reliability without redundancy from a radar monopulse receiver," *Microwaves*, pp. 56–60, December 1967.
19. H. Sakamoto and P. Z. Peebles, Jr., "Conopulse radar," *IEEE Trans*., vol. AES-14, pp. 199–208, January 1978.
20. P. A. Bakut and I. S. Bol'shakov, *Questions of the Statistical Theory of Radar*, vol. II, Moscow: Sovetskoye Radio, 1963, Chaps. 10 and 11. (Translation available from NTIS, AD 645775, June 28, 1966.)
21. M. I. Skolnik, *Introduction to Radar Systems*, 2nd Ed., New York: McGraw-Hill Book Company, 1980.
22. D. K Barton, *Radar Systems Analysis*, Norwood, MA: Artech House, 1977.
23. A. S. Locke, *Guidance, Princton*, NJ: D. Van Nostrand Company, 1955, Chap. 7.
24. D. C. Cross, "Low jitter high performance electronic range tracker," in *IEEE Int. Radar Conf. Rec*., 1975, pp. 408–411.
25. D. L. Malone, "FLYCATCHER," *Nat. Def*., pp. 52–55, January 1984.
26. Hollandse Signaalapparaten B.V. advertisement, *Def. Electron*., vol. 19, p. 67, April 1987.
27. Editor, "Inside the Exocet: Flight of a sea skimmer," *Def. Electron*., vol. 14, pp. 46–48, August 1982.
28. Editor, "Special series: Israeli Avionics-2," *Aviat. Week Space Technol*., pp. 38–49, April 17, 1978.
29. D. C. Cross, D. D. Howard, and J. W. Titus, "Mirror-antenna radar concept," *Microwave J*., vol. 29, pp. 323–335, May 1986.
30. D. D. Howard and D. C. Cross, "Mirror antenna dual-band light weight mirror design," *IEEE Trans*., vol. AP-33, pp. 286–294, March 1985.
31. E. P. Schelonka, "Adaptive control technique for on-axis radar," in *Int. Radar Conf. Rec*., 1975 pp. 396–401.

32. I. D. Olin and F. D. Queen, "Dynamic measurement of radar cross section," *Proc. IEEE*, vol. 53, pp. 954–961, August 1965.
33. J. H. Dunn, D. D. Howard, and A. M. King, "Phenomena of scintillation noise in radar-tracking systems," *Proc. IRE*, vol. 47, pp. 855–863, May 1959.
34. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill Book Company, 1962, Chap. 2.
35. D. K. Barton, *Modern Radar System Analysis*, Norwood, Mass: Artech House, 1988, p. 388.
36. G. Merrill, D. J. Povejsil, R. S. Raven, and P. Waterman, *Airborne Radar*, Boston: Boston Technical Publishers, 1965, pp. 203–207.
37. D. D. Howard, "Radar target angular scintillation in tracking and guidance systems based on echo signal phase front distortion," in *Proc. Nat. Electron. Conf*., vol. 15, October 1959.
38. R. H. Delano, "A theory of target glint or angle scintillation in radar tracking," *Proc. IRE*, vol. 41, pp. 1778–1784, December 1953.
39. R. H. Delano and I. Pfeffer, "The effects of AGC on radar tracking noise," *Proc. IRE*, vol. 44, pp. 801–810, June 1956.
40. J. H. Dunn and D. D. Howard, "The effects of automatic gain control performance on the tracking accuracy of monopulse radar systems," *Proc. IRE*, vol. 47, pp. 430–435, March 1959.
41. D. C. Cross and J. E. Evans, "Target generated range errors," in *IEEE Int. Radar Conf. Rec*., Arlington, VA, April 21–23, 1975, pp. 385–390.
42. D. J. Povejsil, R. S. Raven, and P. Waterman, *Airborne Radar*, Princeton, NJ: D. Van Nostrand Company, 1961, pp. 397–399.
43. R. Hynes and R. E. Gardner, "Doppler spectra of S band and X band signals," *IEEE Trans. Suppl*., vol. AES-3, pp. 356–365, November 1967.
44. A. A. Ausherman, A. Kozma, J. L. Walker, H. M. Jones, and E. C. Poggio, "Development in radar imaging," *IEEE Trans*., vol. AES-20, pp. 363–400, July 1984.
45. G. Dike, R. Wallenberg, and J. Potenza, "Inverse SAR and its application to aircraft classification," in *IEEE Int. Radar Conf. Rec*., pp. 20–25, 1980.
46. D. K. Barton, "The low-angle tracking problem," presented at *IEE Int. Radar Conf*., London, October 23–25, 1973.
47. D. K. Barton and H. R. Ward, *Handbook of Radar Measurement*, Englewood Cliffs, NJ: Prentice-Hall, 1969.
48. D. K. Barton, "Low-angle radar tracking," *Proc. IEEE*, vol. 62, pp. 687–704, June 1974.
49. D. K. Barton, *Radar Resolution and Multipath Effects* in vol. 4 of *Radars*, Norwood, MA: Artech House, 1978.
50. D. D. Howard, J. Nessmith, and S. M. Sherman, "Monopulse tracking error due to multipath: Causes and remedies," in *EASCON Rec*., 1971 pp. 175–182.
51. E. M. T. Jones, "Paraboloid reflector and hyperboloid lens antenna," *IRE Trans*., vol. AP-2, pp. 119–127, July 1954.
52. R. Mitchell et al., "Measurements of performance of MIPIR (Missile Precision Instrumentation Radar Set AN/FPQ-6)," *Final Rept., Navy Contract NOW61-0428d*, RCA, Missile and Surface Radar Division, Moorestown, NJ, December 1964.
53. P. R. Dax, "Accurate tracking of low elevation targets over the sea with a monopulse radar," in *IEE Radar Conf. Publ*. 105, *Radar—Present and Future*, London, October 23–25, 1973, pp. 160–165.
54. D. D. Howard, "Investigation and application of radar techniques for low-altitude target tracking," in *IEE Int. Radar Conf. Rec*., London, October 25–26, 1977.
55. D. D. Howard, "Environmental effects on precision monopulse instrumentation tracking radar at 35 GHz," in *IEEE EASCON '79 Rec*., October 1979.
56. R. J. McAulay and T. P. McGarty, "Maximum-likelihood detection of unresolved targets and multipath," *IEEE Trans*., vol. AES-10, pp. 821–829, November 1974.

57. W. D. White, "Techniques for tracking low-altitude radar targets in the presence of multipath," *IEEE Trans*., vol. AES-10, pp. 835–852, November 1974.
58. P. Z. Peebles, Jr., "Multipath error reduction using multiple target methods," *IEEE Trans*., vol. AES-7, pp. 1123–1130, November 1971.
59. F. E. Nathanson, *Radar Design Principles*, New York: McGraw-Hill Book Company, 1969, p. 37.
60. G. Linde, "Reduction of radar tracking errors with frequency agility," *IEEE Trans*., vol. AES-4, pp. 410–416, May 1968.
61. G. Linde, "A simple approximation formula for glint improvement with frequency agility," *IEEE Trans*., AES-8, pp. 853–855, November 1972.
62. D. K. Barton, *Frequency Agility and Diversity*, in Vol. 6 of *Radars*, Norwood, MA: Artech House, 1977.

Chapter 10
The Radar Transmitter

Thomas A. Weil and Merrill Skolnik

10.1 INTRODUCTION

Role of the Transmitter in Radar. If a radar systems designer could ask for anything he or she wanted in a radar transmitter, that wish might be something like the following:

> Provide the necessary transmitted energy with the needed average and peak power, as well as the required stability and low noise for good doppler processing; operate with high efficiency; have wide bandwidth and be easily tunable; be readily modulated in amplitude, frequency, or phase as necessary; have high reliability and long life; require minimum maintenance; have no dangerous X-ray emissions; require no personnel to operate; be of an affordable price; and be of reasonable size and weight for the desired application.

Of course, not all of these desirable attributes can be achieved in any given radar transmitter application. Compromises need to be made depending on the application.

The radar equation for a surveillance radar (one that has to cover a fixed volume of space on a regular basis) indicates that the maximum range of such a radar is proportional to $(P_{av}A)^{1/4}$, where P_{av} is the average power of the transmitter and A is the area of the antenna aperture.[1] A fundamental measure of the radar range performance is therefore the *power-aperture product*. One can obtain a long range by having a large antenna, a large transmitter, or a combination of both. It is not usual, under most circumstances, for a radar to have a huge, costly antenna and a small, inexpensive transmitter, or vice versa. There has to be a reasonable balance between these two major subsystems of a radar. (In a surveillance radar, straightforward calculus will show that under simple assumptions the minimum total cost of a radar occurs when the cost of the transmitter equals the cost of the antenna; but this is true only when there are no criteria, other than minimum cost, that have to be satisfied—and there are usually other criteria that need to be considered.)

As is well known in radar, the doppler effect is extensively used to detect moving targets in the presence of large clutter echoes. It is the basis for several of the chapters in this Handbook. Some transmitter types, however, are far better than others when the radar has to employ the doppler-shifted signal to detect moving targets in the midst of heavy clutter echoes.

Examination of the basic radar equation for detection of targets at long range indicates that the average power is far more important than the peak power as a measure

of the radar's capability. A vacuum tube* with a given average power can usually be designed to handle the high voltages associated with a large peak power without breakdown. A solid-state transmitter cannot. In the past, average powers of radar transmitters have been from a small fraction of a watt to the order of a megawatt.

Types of Radar Transmitters. The very first "radars," such as those used by Heinrich Hertz (the first radar scientist) in the late 1880s and the shipboard radar invented by Christian Hulsmeyer (the first radar engineer) in the early 1900s used the *spark gap* as the transmitter. It was a very poor transmitter, but that is not unusual in the early days of a new and different development. The DeForest *grid-controlled vacuum tube* (triode) was invented shortly thereafter, and by the early 1930s, it had been well developed to where it was successfully and extensively used in those countries that built the first VHF and UHF radars used for air defense early in World War II. Some UHF radars effectively used the grid-controlled vacuum tube well into the early part of the current century. It has been a very competitive power source for UHF radar applications. The drawback of the grid-controlled vacuum tube is that transit-time effects limit its application at microwave frequencies, but variants of grid-controlled vacuum tubes have been successfully used up to about 1000 MHz.

The barrier of transit time effects was overcome with the invention of the microwave cavity *magnetron* early in World War II in 1940 by the United Kingdom (UK). The introduction of the magnetron allowed high-power radar to be successfully developed for use at the higher frequencies where smaller size antennas could be used. (It is of interest to note that the Japanese invented the magnetron before the UK did, and Soviet Union engineers published a paper describing their magnetron in the March 1944 issue of *Proceedings of the Institute of Radio Engineers,* now the *Proceedings of the IEEE*; but the wartime chaos in the military developments of both the Soviet Union and Japan was such that these inventions were not fully exploited by these two countries during WWII.) The invention of the magnetron was important because it allowed radar to be developed for use at microwave frequencies rather than be limited to VHF and UHF frequencies. It caught the German electronic countermeasures effort off guard because they had no idea that the United Kingdom and the United States could produce microwave radar. The success of the magnetron was a large factor in the effective application of military radar by the United States and the United Kingdom during World War II.

The magnetron is an example of what is called a *crossed-field tube* in that it employs a magnetic field and an electric field that are orthogonal to one another. The magnetron is an oscillator but the grid-controlled vacuum tube could be operated as either an oscillator or an amplifier. The other electronic tubes mentioned in this chapter are usually operated as amplifiers. Amplifiers at microwave frequencies have generally produced higher powers than oscillators; but, probably more important, they allow the use of stable, modulated waveforms needed for waveforms in radars that depend on the use of pulse compression and for the doppler effect to detect moving targets in clutter.

The microwave *klystron* amplifier was invented before the magnetron and was described in a paper in the May 1939 issue of the *Journal of Applied Physics*. For the most part, the klystron amplifier was ignored during WWII and didn't attract

---

* In the United States, the device that generates the RF power is called a *tube,* but in the United Kingdom, it is called a *valve*. A book on microwave power sources[4] suggests that these devices be called *microwave vacuum electronic devices (MVED)*. In this chapter, however, the name *tube* will be retained.

the attention of radar engineers until the announcement in a paper in the November 1953 *Proceedings of the IEEE* by Stanford University engineers of the development of an S-band multicavity klystron capable of 20 MW peak power and 2.4 kW average power for use in a linear accelerator. This was a great accomplishment in its time. The high power, high efficiency, good stability, and wide bandwidth (at high power) of the microwave klystron amplifier have caused some radar design engineers to say that the klystron should be the first microwave power source to consider when designing a new high-performance radar. (There was, at one time, a single-cavity klystron oscillator called the *reflex klystron* that was of low power and mainly used as a receiver local oscillator, but it has generally been replaced by solid-state devices and is not discussed further in this chapter.)

The klystron is an example of *a linear-beam tube* because the direction of the dc electric field that accelerates the electron beam coincides with the axis of the magnetic field that focuses and confines the beam. It generates a highly concentrated high-energy linear beam of electrons that interacts with the microwave structure (two or more microwave cavities) to achieve amplification. Another example of the linear-beam tube is the *traveling wave tube* (TWT) amplifier. It generally can do almost what a klystron can do, but it is capable of very wide bandwidth at low powers, which the klystron is not. The TWT usually has slightly less gain than a klystron and less stability. It should be noted, however, that as the power of a TWT increases, its bandwidth decreases; and as the power of a klystron amplifier increases, its bandwidth increases. Thus, at the high powers needed for many radar applications, the bandwidths of these two types of linear-beam tubes are approximately comparable.

There also have been *hybrids* of the klystron and the TWT that have been of interest for radar applications, and which have interesting characteristics.

The *crossed-field amplifier*, like the magnetron, is a crossed-field tube that employs a magnetic field orthogonal to the electric field. It is capable of wide bandwidth and generally is of smaller size and does not require the very high voltages of the linear-beam tube. Although it has some advantages not found in other tubes, it has lower gain than linear-beam amplifiers (so multiple stages of amplification are required), and its noise level is higher than the linear-beam tube, which makes it less capable for detecting moving targets in clutter.

The *gyrotron*, which can be either an oscillator or an amplifier, is an RF* power source that can produce very high power at millimeter wavelengths. Conventional microwave power sources utilize resonant structures in which the phase velocity of the electromagnetic field propagating along the RF structure is slowed so as to be close to the electron beam velocity. Thus, they are known as *slow-wave tubes*. The characteristic size of the increments of the RF structure of slow-wave tubes is typically a fraction of a wavelength. They become smaller as the frequency is increased (wavelength is decreased). Smaller size means that a tube cannot dissipate heat as well as a larger tube, so that the power capability of microwave power tubes decreases approximately inversely as the square of the frequency. The gyrotron, on the other hand, does not have this type of frequency dependence since it uses what is called a *fast-wave* structure. This is usually a smooth waveguide or a large resonator. No attempt is made to reduce the propagation velocity of the electromagnetic wave within this structure. The electron beam is not close to the RF structure so it is not as limited in size as are the structures of

* In electrical engineering, RF stands for radio frequency, but in radar, it is often used to mean the radar frequency.

slow-wave tubes. The larger size of the fast-wave tube means that it can handle larger power at the higher frequencies. The gyrotron has mainly been of significance for high-power applications at millimeter-wave frequencies.

The *solid-state transistor amplifier* has been of interest for radar applications. Such interest in solid state has also been due, in part, to its completely replacing vacuum tubes in receiver and computer applications. The solid-state transmitter is discussed in Chapter 11, and a brief comparison of it with the vacuum tube transmitter is given at the end of this chapter. The chief advantages of a solid-state radar transmitter are that it can operate with wide bandwidths; it has the potential for long life; and it is favored by some buyers of radar. It cannot, however, employ high peak-power waveforms. The limitation on peak power in a solid-state transmitter results in compromises that have to be made in the design of the overall radar system.

Oscillator Versus Amplifier. The power amplifier is often preferred over the power oscillator as the transmitter power source in high power, high performance radar systems. In an amplifier, the signal to be transmitted is precisely generated at a low power level and is then amplified to achieve the required power to be radiated from the antenna. Amplifiers have the advantage of being able to provide stable waveforms, coded or frequency modulated pulse compression waveforms, frequency agility, as well as combining and arraying.

The magnetron is an oscillator that has less flexibility and is usually noisier than a linear-beam amplifier. Each time a pulse is transmitted, its phase is different from the phase of previous pulses. That is, its phase is random from pulse to pulse. To detect the doppler frequency shift for MTI processing, the phase cannot change in a random manner at the receiver from pulse to pulse. This limitation is overcome by taking a sample of the random phase of each transmitted pulse and using it to reset the phase of the local oscillator in the receiver to match the phase of the transmitted signal. This is sometimes called *coherent on receive*. Generally, the MTI improvement factor that can be obtained with a magnetron is not as good as can be obtained with a linear-beam amplifier.

In the past, there might have been debate as to whether to use an oscillator or an amplifier for a high-performance radar transmitter. There is usually no question that the amplifier is usually the preferred choice, except in situations where the low cost of a magnetron transmitter is more important than the lower MTI improvement factor it provides compared to a linear-beam transmitter. The magnetron oscillator, however, has been used in some short or medium range radars and in the widely popular civil marine radar (Chapter 22), which requires only a small power transmitter and has no need for MTI capability.

10.2 LINEAR-BEAM AMPLIFIERS[2]

The klystron, TWT, and hybrids of the two have been important sources of RF power for many successful radar systems. The electrons emitted from the cathode are formed into a long cylindrical beam that receives the full potential energy of the electric field before the beam enters the microwave interaction region. The electron beam, generated by an "electron gun," essentially flows in a straight line in linear-beam tubes to interact with a microwave circuit to produce amplification of an input signal. The major difference among the several types of linear-beam tubes is the type of microwave circuit employed

and the nature of the interaction that produces amplification. Transit-time effects, which limit the high frequency performance of grid-controlled tubes, are taken advantage of in linear-beam tubes to velocity modulate the uniform electron beam to create bunches of electrons from which RF energy can be extracted at the output of the tube. Linear-beam tubes as amplifiers can produce high power with good efficiency and high gain and with wide bandwidth. They have been capable of producing average powers of a megawatt, as well as average powers of many kilowatts in a size suitable for use in a military fighter/attack aircraft.

Klystron. The klystron amplifier has been an important RF power source for many radar applications. As mentioned, it is capable of high average and high peak power, high gain, good efficiency, stable operation, low interpulse noise, large bandwidth at high power; and being an amplifier, it can work well with the frequency and phase modulated waveforms needed for pulse compression. Klystrons have operated at frequencies from UHF to millimeter wavelengths and have found use in such diverse applications as airport surveillance radars where the average powers might be more than one kilowatt, in airborne military aircraft where the average power might be of the order of 10 kW or more, and in long range detection of intercontinental ballistic missiles where the average power per tube can be greater than 100 kW.

Figure 10.1 depicts the principal parts of a three-cavity klystron. At the left is the *heater* that heats the cathode and the *cathode* that emits a stream of electrons that are collimated into a narrow cylindrical beam of high electron density. The *electron gun* that generates the beam consists of the *cathode,* the *modulating anode* (also called the *beam control grid*), and the *anode*. The modulating anode provides the means to pulse the electron beam on and off. The *RF cavities* are the microwave equivalent of a resonant circuit. Electrons are not intentionally collected at the anode (as they are in grid-controlled tubes and crossed-field tubes), but in the *collector,* shown at the right-hand side of the illustration, after the electron beam has given up its RF energy at the output cavity. A low-power signal is applied to the input of the first cavity and appears at the *interaction gap*. Those electrons in the beam that arrive at the first interaction gap when the input signal voltage is at a maximum (the positive peak of the sine wave) are speeded up compared to those electrons that arrive at the gap when the input signal voltage is a minimum (the negative trough of the sine wave).

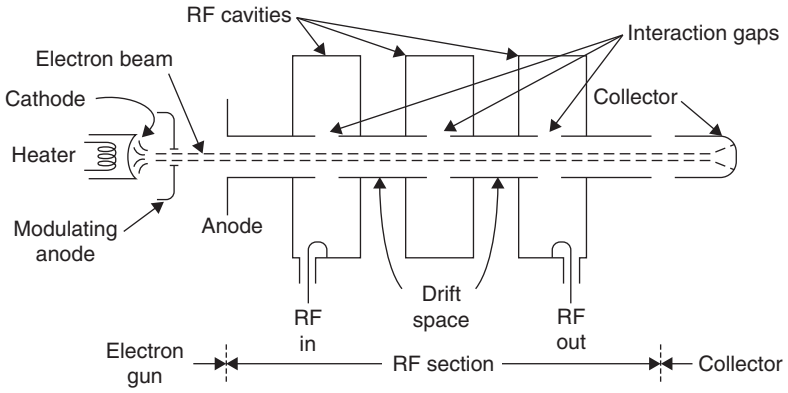


FIGURE 10.1 Representation of the principal parts of a three-cavity klystron amplifier

In the first *drift space,* those electrons speeded up during the peak of one cycle catch up to those that were slowed down during the minimum of the previous RF cycle. The result is that the electrons become "bunched" periodically. This bunching can be thought of as producing a modulation of the density of electrons. The bunches pass through the interaction space of the second cavity, which reinforces the *density modulation* to enhance the bunching. This process of impressing a time variation in velocity that results in bunching of the electrons of an initially uniform electron beam is called *velocity modulation*. Three or more RF cavities might be used. The interaction gap of the output cavity is placed at the point of maximum bunching so that the RF power can be extracted from the density modulated electron beam by a coupling loop in a lower power tube or by a waveguide (not shown) in a high power tube. In essence, the dc energy of the electron beam at the first cavity is converted to RF energy at the output cavity by the velocity modulation process. The larger the number of cavities the greater can be the gain of the klystron. The gain of a four-cavity klystron can be more than 60 dB, depending on the bandwidth. After the bunched electrons deliver their RF power to the output, the spent electrons are removed by the *collector* electrode.

An axial magnetic field is employed to counteract the mutual repulsion of the electrons that form the electron beam. The magnetic field confines the electrons to a relatively long, thin beam and prevents them from dispersing. It can be generated by a long solenoid that has iron shielding around its outside diameter, or by a lighter weight periodic-permanent-magnetic (PPM) system that consists of a series of magnetic lenses.

A multicavity klystron can have its bandwidth increased by stagger tuning the cavities, similar to the manner in which stagger tuning is done in an IF amplifier of a superheterodyne receiver to obtain broader bandwidth. It is more complicated, however, to do this in the klystron since the velocity modulation that appears at each interaction gap contributes a component to the exciting current at the succeeding gaps, something that does not occur in an IF amplifier. The early VA-87 four-cavity S-band klystron had a 20 MHz bandwidth and a gain of 61 dB when its four cavities were synchronously tuned, but when stagger-tuned, its bandwidth was 77 MHz (2.8%) and a gain of 44 dB.[3]

Theory shows that the bandwidth of a klystron can be significantly increased by increasing its current and thus its power. A 10 MW peak power klystron, for example, can have an 8% bandwidth as compared to a 200 kW tube, which might have a 2% bandwidth, and a 1 kW tube having only a 0.5% bandwidth. High-power multicavity klystrons can be designed with bandwidths as large as 10 to 12%. The klystron is sometimes thought of as having a narrow bandwidth and the traveling wave tube is thought of as having wide bandwidth; but at the high power levels needed for long range radars, their bandwidths are comparable. It is sometimes unfortunate that this fact is not always understood.[4]

The klystron, as an example of a linear-beam tube, is capable of high power because the generation of the electron beam, its interaction with the electromagnetic field, and the collection of the spent electrons are performed in separate parts of the tube where the generated heat can be dissipated effectively.

The klystron and other linear-beam tubes can have long life. Gilmour[5] reports that the mean time between failures (MTBF) of 11 different applications of klystrons in radar systems varied from 75,000 hours to 5,000 hours, with an average of 37,000 hours for all 11 applications. (There are 8,760 hours in a year.) The VA-842

high-power klystron tube used in the original Ballistic Missile Early Warning System had a demonstrated life in excess of 50,000 hours. Symons[6] reports that one of the BMEWS tubes that he designed was still operating after 240,000 hours when the radar in Greenland was replaced by the solid-state Pave Paws radar. The VA-812E was also a high-power UHF klystron that had a 20 MW peak power, 25 MHz 1 dB bandwidth, 40 dB gain, and an average power of 300 kW at a duty cycle of 0.015 and a 40 μs pulse width.

The VA-87E (originally developed by Varian Associates, Inc) is a 6-cavity S-band pulse klystron that operates from 2.7 to 2.9 GHz, produces a peak power of from 1 to 2 MW, an average power up to 3.5 kW, has a gain of about 50 dB, an efficiency between 45 and 50%, and a 1 dB bandwidth of 39 MHz. It has demonstrated a mean-time-between-failures of 72,000 hours. It was used in the ASR-9 airport surveillance radar and in the WSR-88D Nexrad doppler weather radar (where its operating band is from 2.7 to 3.0 GHz), as well as in other radars.

Very high-power klystrons are employed in linear accelerators such as found at the Stanford Linear Accelerator Center.[7] Klystrons for this application, for example, might have 75 MW peak power with 50% efficiency using solenoid magnets; or 60–75 MW peak power with 60% efficiency using periodic permanent magnets.

Improvements to the klystron as a radar power source are discussed later in the subsection on hybrids, of which the clustered-cavity klystron is a good example of what can be provided in the way of high power and wide bandwidth.

Multiple-Beam Klystron (MBK). The klystron amplifier is an important tube for high-power, high-performance radar applications. However, it requires a large voltage when high power is needed. High voltage results in greater size and the need for shielding from the X-rays that are generated. In a klystron, the power available for conversion of the dc power of the electron beam to RF power of an electromagnetic wave is given by the product of the beam current and the beam voltage.[8] Although klystrons have been operated at very high voltages, it is usually preferred to operate it at a lower voltage, if possible, because lower voltages generally make the power supplies simpler, lighter, and more reliable.[9] A reduction of beam voltage in a conventional klystron means an increase of the beam current in order to maintain the same power. An increase in beam current, however, results in an increase in the current density where space charge effects may not be negligible, and the repulsive forces that occur among the electrons are increased and can cause the electron density bunches to lose coherence. The result is a decrease in efficiency. Higher current densities also require stronger magnets to keep the electron beam confined, leading to larger volume and weight. Thus, simply lowering the beam voltage and increasing the current density does not usually provide a net advantage.

The limitations of lower beam voltage, however, can be overcome by separating the single electron beam into a number of smaller beams, called *beamlets,* so that each of the beamlets has a low enough current density to avoid the undesirable repulsive effects of a high current-density beam. According to Nusinovich et al.,[9] each beamlet is transported down its own individual drift channel (a metallic-walled tube), parallel to, but isolated from, the other beamlets. They are allowed to interact only over the small axial extent of the cavity gap. After passing by the cavity gap, the beamlets reenter their individual drift channels and propagate in isolation from one another. Such klystrons are called *multiple-beam klystrons,* or *MBK*. It has been said[10] that the number of beamlets in an MBK might be from 6 to 60.

The chief motivation for the multiple-beam klystron is the efficient generation of high RF power at a lower voltage than in a conventional klystron. Because of its lower voltage (two or three times lower), an MBK can be more compact, have a lower magnet weight (up to ten times less), be of lower weight and volume, generate less X-rays, have high electronic efficiency (up to 65 percent), and have the potential for a higher instantaneous bandwidth than does a conventional klystron.[9] The lower voltages at which they operate result in power supplies that can be simpler, lighter, cheaper, and more reliable. The MBK can have a high output-power-to-weight ratio that might be two to three times greater than that of an equivalent single-beam klystron. They can also have reduced noise and less phase sensitivity to deviations in voltage, which aids in the detection of low cross-section moving targets in clutter. Compared to a crossed-field amplifier, they have a larger dynamic range. Compared to a TWT, they are capable of higher peak and average powers, and they are less sensitive to vibrations.

The Russian company known as Federal State Unitary Enterprise, RPC Istok[11] (usually shortened to Istok) has been productive in the development of MBKs for radar. At X band, they report an MBK with 24 beamlets producing 200 kW peak power, 17 kW average power, 6% bandwidth, an anode voltage of 26 kV, with a magnet weighing 16 kg. At S band, one of their tubes with 36 beams had a peak power of 600 kW, 12 kW average power, operated with 31 kV anode voltage, had a 6.5% bandwidth, and a weight without magnet of 25 kg.

An apparent extension of the multiple beam concept is to employ a sheet electron beam, which is as thin and as wide and has as much current as can be achieved consistent with other constraints. It has been considered for very high-power klystrons (150 MW peak power and periodic permanent magnetic, or PPM, focusing) designed for a very large linear accelerator.[12] It is claimed that the beam current density and the focusing magnetic field can be reduced, be made with fewer parts, might be more reliable, and can have lower acquisition and operating costs. One possible disadvantage is that sheet-beam klystrons might not be wideband.

The principle of the multiple-beam linear-beam tube has also been considered for the traveling wave tube,[9] but it is not obvious whether it has any significant advantages over the MBK.

Traveling Wave Tube (TWT). The TWT linear-beam tube is similar to the klystron in that the cathode, RF circuit, and collector are all separate from one another. There is, however, continuous interaction of the electron beam and the RF field over the entire length of the microwave propagating structure of the TWT as compared to the klystron where the interaction occurs only at the gaps of a relatively few resonant cavities. The TWT was originally conceived with a helix as the slow-wave RF structure, as in the illustration shown in Figure 10.2. The electron beam is similar to

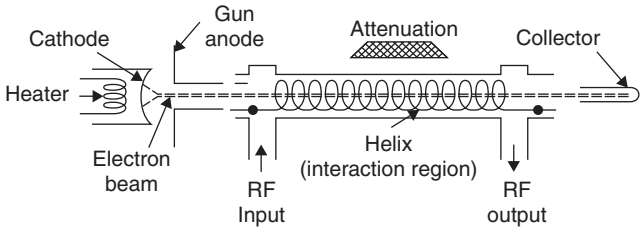


FIGURE 10.2 Representation of the principle parts of a traveling wave tube, showing a helix slow-wave circuit shown for simplicity

that of the klystron, and they both use the process of velocity modulation to cause the electron beam to be periodically bunched (density modulation). The electron beam passes through the RF interaction circuit. In the example shown in Figure 10.2, where a helix is shown as the slow-wave structure, the RF signal is slowed down by the helix so that its forward velocity is very nearly equal to that of the velocity of the electron beam. This near match of velocities is what causes the cumulative interaction that transfers dc energy from the electron beam to amplify the electromagnetic wave propagating on the helix. After delivering its energy to the RF field, the spent electrons are removed by the collector, usually a multistage depressed collector. An axial magnetic field keeps the electron beam from dispersing as it travels down the tube, just as in the klystron.

The helix TWT is capable of bandwidth in excess of an octave (2 to 1), which is much higher than other radar tubes. The TWT is often thought of as a very broadband tube, but the broad bandwidth of a helix TWT is not of great significance in radar applications since the helix TWT is limited in its peak power to a few kW. This means that the helix TWT is best used in CW or high duty cycle radar applications, if at all. Also, such broad bandwidths are seldom available for radar applications because of regulatory restrictions in the use of the electromagnetic spectrum. To achieve high power in a TWT, other types of slow-wave structures have to be employed, and these usually provide less bandwidth than the helix. Such structures are the coupled cavity[13] (of which the cloverleaf is an example), ring-bar, and the so-called ladder network. The bandwidth of a coupled-cavity TWT can be 10 to 15%. The ring-bar has broader bandwidth and higher efficiency than the coupled-cavity circuit, but it is not capable of as high a power as the coupled cavity. Thus, the bandwidth of a TWT decreases as its power increases. On the other hand, as mentioned previously, the bandwidth of a klystron increases with increasing power, so that the bandwidths of the TWT and the klystron are generally comparable for many high-power radar applications.

Both forward and backward waves may propagate along the microwave structure of a TWT, leading to a possibility of backward-wave oscillations. In Figure 10.2, attenuation is shown along the helix so as to prevent oscillation due to reflections at the output and the input of the structure. The attenuation may be distributed or lumped, but it is usually found in the middle third of the tube. Although oscillation can be prevented by distributing loss along the structure, it results in lower efficiency—something unattractive in high-power tubes. Instead, oscillations may be prevented by the use of discontinuities called severs, with one sever for every 15 to 30 dB of tube gain. At each sever, the power traveling in the reverse direction is dissipated in the sever loads without seriously affecting the power traveling in the forward direction. The sever loads may be placed external to the tube to reduce dissipation within the RF structure itself. The efficiency of a TWT is usually less than that of a klystron because of the loss due to the attenuation of the severs, as well as by the presence of relatively high RF power over an appreciable part of the entire structure. A technique for improving the efficiency of high-power TWTs is called *velocity tapering*. This technique consists of tapering the length of the last few sections of the slow-wave circuit to take into account the slowing down of the beam as the energy is extracted from it. Velocity tapering permits extracting more of the energy from the beam and significantly improves the power-bandwidth performance of the tube.[3] Nevertheless, high-power TWTs generally show an appreciable falloff of power output toward the band edges, so that the rated bandwidth depends very much on how much power falloff can be tolerated by the system.

If a TWT using a coupled-cavity circuit is cathode-pulsed (Section 10.7), there is an instant during the rise and fall of voltage when the beam velocity becomes synchronous with the cutoff frequency (the so-called π mode) of the microwave circuit, and the tube can generate oscillations. These oscillations at the leading and trailing edges of the RF output pulse have a characteristic appearance on a power-time presentation that has given them the name *rabbit ears*. Only in rare cases has it been possible to suppress these oscillations completely. However, since this particular oscillation depends on electron velocity, which in turn depends on beam voltage, the problem is avoided by the use of mod-anode or grid pulsing (Section 10.7). In this case, it is only necessary to be sure not to let the modulator begin pulsing the beam current during turn-on of the high-voltage power supply until the voltage is safely above the oscillation range, which is typically somewhere between 60 and 80% of full operating voltage.

A modification of the helix slow-wave structure is the ring-bar circuit, which can be used if the peak power is less than 100 to 200 kW. The Raytheon QKW-1671A is an example, with a peak power of 160 kW, 0.036 duty cycle, and 50 dB gain. It operates at L band from 1215 to 1400 MHz. The U.S. Air Force Cobra Dane, operating from 1175 to 1375 MHz, is a long range radar located in the Aleutian Islands that uses 96 ring-bar TWTs (QKW-1723), each with a peak power of 175 kW and average power of 10.5 kW.

The S band VA-125A TWT amplifier employs a cloverleaf coupled cavity microwave slow-wave structure. With the use of liquid cooling, it is capable of 3 MW peak power over a 300 MHz bandwidth. Its duty cycle is 0.002, with a gain of 33 dB, and a 2 μs pulse width. This TWT was originally designed to be used interchangeably with the popular VA-87 klystron, except that the VA-125A TWT has a wider bandwidth than the VA-87 klystron. It also requires a greater power input signal because of its lower gain than the klystron. When high power is required, the klystron tends to be preferred over the TWT since it doesn't experience the stability problems of the TWT.

Gilmour[14] gives the mean time between failures (MTBF) for nine different types of coupled-cavity TWTs as varying from 17,800 hours to 2,200 hours, with an average of 7,000 hours for all nine classes of tubes. (He also says that TWTs for space application, which are of lower power than radar TWTs, have MTBFs of the order of one million hours.)

Depressed Collector. The efficiency of a TWT or a klystron can be improved by the use of a so-called *depressed collector*.[15,16] With a single collector, a significant fraction of the power input to the tube is dissipated as heat in the collector. If the voltage on the collector is reduced (depressed) below the body voltage, the velocity of the electrons striking the collector is reduced and so is the heat generated in the collector. Thus, the collector recovers some of the power in the spent electron beam. The use of multiple depressed collectors at intermediate voltages, rather than a single collector, allows catching each spent electron at a voltage near optimum. Up to ten collector sections have been used in some communication tubes, but three sections are more typical for high-power TWTs for radar systems. The several different voltages needed for the depressed collectors add complexity to the high-power voltage supply, but these voltages need not be as well regulated as the main-beam voltage. It is usually easier to design a depressed collector for a TWT than for a klystron since the spent electron beam of a TWT might have a 20% spread in velocity, but the klystron might have a velocity spread of almost 100%.[17] Because the efficiency in a conventional TWT is usually lower than that of a klystron, the increase in efficiency in the TWT provided by a depressed collector has a greater relative effect than with a klystron.

Variants of the Klystron and the TWT. It was mentioned that the bandwidth of a klystron increases as its power is increased. The bandwidth can also be increased by combining the best features of the klystron and the traveling wave tube to obtain better bandwidth, efficiency, and gain flatness than either the conventional klystron or the TWT. In these tubes, the basic structure is that of a klystron, but instead of a number of single cavities being used, the single cavity is replaced by a more complex multiple cavity. Three such variants are the *Twystron, extended interaction klystron,* and the *clustered cavity klystron*. Most high-performance radar klystrons tend to employ the more intricate cavity structure because of the better performance it provides.

Comparison of Various Linear-Beam Tube Structures. Figure 10.3 illustrates the basic structure of the RF circuits that characterize the various types of linear-beam tubes.

Twystron. The bandwidth of a conventional klystron is limited primarily by the bandwidth of the output resonant cavity. If a coupled cavity slow-wave circuit, as is used in the TWT, is substituted for the output resonant cavity of a klystron (Figure 10.3*d*), the bandwidth of the klystron can be increased significantly, and there is a slight increase in efficiency. This requires that the intermediate cavities and the input cavity of such a tube be stagger-tuned to accommodate the increased bandwidth offered by the output circuit. Because this type of tube is part klystron and part TWT, it was named *Twystron*. The VA-145 S-band Twystron has a bandwidth of 14% with a 35% efficiency, 41 dB gain at midband, peak power of 3.5 MW, and average power of 7 kW.[19]

Extended Interaction Klystron (EIK). In the EIK, the single-gap resonant cavities of the klystron are replaced by a resonated slow-wave TWT-like circuit that contains two or more interaction gaps (Figure 10.3*c*). Such cavities can be used for the prior cavities as well as the output cavity. This allows wider bandwidth and greater power than the conventional klystron amplifier. Staprans et al.[2,18] state that the high-power VA-812C EIK operated over a frequency range from 400 to 450 MHz (12% bandwidth), with a peak power of 8 MW and an average power of 30 kW, and an efficiency of 40%.

EIK devices have been of interest for millimeter-wave applications. According to the Communications & Power Industries (CPI) brochure, its VKB 2475 millimeter-wave EIK when operating at a center frequency of 94.5 GHz with a 1 GHz bandwidth, has a peak power of 1.2 kW, average power of 150 W, about 10% duty cycle, gain of 47 dB, pulse width of 20 μs, and is liquid cooled. It has much lower power than a gyroklystron (discussed later) at this frequency, but its 18 cm by 10 cm (diameter) size is considerably smaller and it costs considerably less than a gyroklystron.

A similar EIK, also built by CPI, has been used in the NASA CloudSat spaceborne radar that provides the vertical profile of clouds for understanding cloud effects on both the weather and climate.[20] It operates at a center frequency of 94 GHz with a 250 MHz bandwidth, 1.5 kW peak power, 3.3 μs pulse duration, 4300 Hz prf, and an efficiency of 32%. It is conduction cooled. Each cavity is a short piece of resonant slow-wave structure based on ladder geometry. The tube weighs 6.2 kg and can operate over a temperature range from −15 to +60°C. It was predicted that this EIK would be capable of two years of continuous operation with a 92% confidence level. Two EIKs were employed in CloudSat (one prime, one redundant) so as to achieve a 99% confidence of meeting the two-year life requirement.

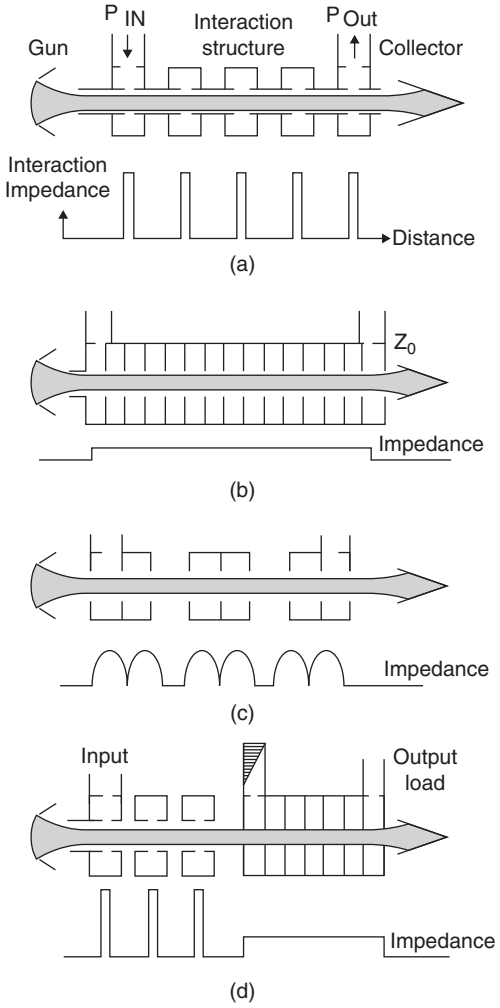


FIGURE 10.3 Basic structure of several types of linear beam tubes: (*a*) klystron, (*b*) coupled-cavity TWT, (*c*) extended interaction klystron, and (*d*) Twystron (*after A. Staprans et al.*[18] *© IEEE 1973*)

Clustered-Cavity Klystrons. This is a good example of the technique of grouping cavities to improve the operation of a klystron. The individual intermediate cavities of a multicavity klystron are each replaced by a cluster of two or three artificially loaded low-Q cavities with Qs of one half to one third of the single cavity they replace.[21] Figure 10.4 compares schematically the basic difference between the conventional stagger-tuned klystron and the clustered-cavity klystron. It has been said for a given

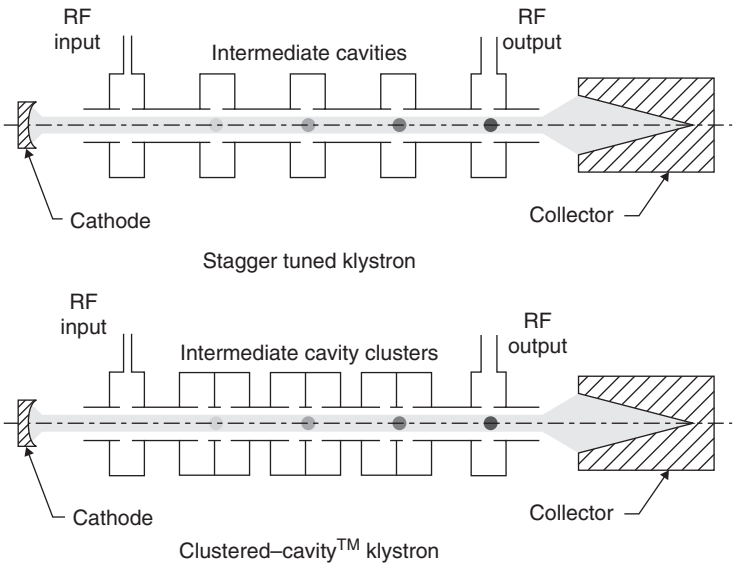


FIGURE 10.4 Comparison of the structures of a conventional stagger-tuned klystron (top) and clustered-cavity klystron (bottom)[16] (*Courtesy of the IEEE*)

gain-bandwidth product, this form of structure can produce a tube of much shorter length so that there can be substantial savings in magnet weight and power. Symons,[22] the inventor of the clustered-cavity klystron, states that one of these wide bandwidth tubes can be used to replace the two narrower-band klystrons in the AWACS radar. When each of the two narrow-band tubes is replaced by a wideband clustered-cavity tube, redundant operation can be provided with higher reliability and without a large weight penalty because either of these clustered-cavity klystrons provide full operational capability similar to the redundancy commonly employed in FAA air-traffic control radars.

Microwave Power Module (MPM).[23,24] A novel variant of the linear-beam tube is the *microwave power module,* which is an amplifier that employs a solid-state microwave integrated-circuit amplifier to drive a moderate-power helix traveling wave tube, along with an integrated circuit power conditioner, all in a lightweight package. It can provide high efficiency, wide instantaneous bandwidth, low noise, and average power levels from several tens to several hundreds of watts. It is said to be smaller and lighter than comparable TWT and solid-state power sources, and is capable of operating at high ambient temperatures. The gain of an MPM might be nominally 50 dB and is divided between the solid-state driver and the TWT power booster in the ratios from 20/30 to 30/20. The MPM seems best suited for the higher microwave frequencies, perhaps from 2 to 40 GHz.

A serious constraint of the MPM for radar applications is that the helix TWT limits its use to CW or high duty cycle transmissions (preferably greater than 50%). It also is of relatively modest power for many radar applications.

10.3 MAGNETRON

Unlike linear-beam tubes that are normally operated as amplifiers, the magnetron is an oscillator. As example of a commonly used early magnetron was the 5J26, an L-band tube that was mechanically tunable from 1250 to1350 MHz. It could operate with a peak power of 500 kW, a 1 μs pulse width, and a pulse repetition frequency of 1000 Hz that provided an average power of 500 W. Its 40% efficiency was typical for magnetrons at that time. The compact size and efficient operation of the magnetron at microwave frequencies allowed radars in World War II to be small enough to fly in military aircraft and to be mobile for ground warfare. Magnetrons, however, seem to be limited to about a few kilowatts of average power, which can restrict their utility. They also have limitations in stability and, therefore, in the MTI improvement factor they can achieve, and they often have a shorter life than linear-beam tubes.

Because the magnetron is an oscillator rather than an amplifier, the starting phase of each pulse is random from pulse to pulse. This random change of phase can be accommodated in a MTI radar receiver by use of a coherent oscillator (coho) as the reference signal in the phase-detector stage of the receiver. On each pulse, the phase of the magnetron pulse sets the phase of the coho. In this manner, the received signal appears to be coherent pulse to pulse. This is sometimes called *coherent on receive*. The MTI improvement factor obtained with a magnetron and a coherent-on-receive operation usually is not as good as can be obtained with an MTI system that uses a power amplifier as the transmitter.

Automatic frequency control (AFC) is often employed to keep the receiver tuned to the frequency of the transmitter since the magnetron frequency can slowly drift with changes in ambient temperature and self-heating. The AFC can be applied to the magnetron itself to keep it operating on its assigned frequency, within the limits of the accuracy of the tuning mechanism.

A magnetron can be mechanically changed in frequency over a 5 to 10% frequency range and, in some cases, as much as 25%. Rapid mechanical tuning can be achieved with a slotted disk suspended above the anode cavities. When rotated, it alternately provides inductive and capacitive loading of the cavities to raise and lower the frequency. Such a rotary-tuned magnetron can provide very fast tuning rates. For example, at a rotation rate of 1800 rpm, a magnetron with 10 cavities can tune across a band 300 times per second.

Coaxial Magnetron.[25] A significant improvement in the power, efficiency, stability, and life of the original form of the magnetron occurred with the introduction of the coaxial magnetron. The key difference is the incorporation of a stabilizing cavity surrounding the conventional magnetron cavities, with the stabilizing cavity coupled to the magnetron cavities so as to provide better stabilization. The frequency of a coaxial magnetron can be changed by mechanically moving one of the end plates, called a *tuning piston,* of the stabilizing cavity. The tuning piston can be positioned mechanically from the outside of the vacuum by means of a vacuum bellows.

In the coaxial magnetron, the output of every other resonant cavity is coupled to the stabilizing cavity that surrounds the anode structure. The output power is then coupled from the stabilizing cavity.

π Mode of Operation. A magnetron, whether conventional or coaxial, can oscillate at a number of different, closely spaced frequencies due to various possible configurations of the RF field that can exist between the cathode and the resonant cavities.

These different RF field configurations, along with coupling among the cavity resonators of the magnetron, result in different modes of oscillation. The magnetron can shift, almost unpredictably, from one mode to another (which means the frequency shifts unpredictably) as the voltage changes or as the input impedance that the magnetron experiences changes. The shift from one mode to another, often called *moding*, is especially bad since it can occur when the radar antenna scans and views different environments. It is important to avoid moding.

The preferred magnetron mode of operation is the so-called π *mode* that occurs when the RF field configuration is such that the RF phase alternates 180° (π radians) between adjacent cavities. The advantage of the π mode is that its frequency can be more readily separated from the frequencies of the other possible modes. (An N-cavity magnetron can have $N/2$ possible modes. The π mode oscillates at a single frequency, but the other modes can oscillate at two different frequencies so that the magnetron can oscillate at $N - 1$ different frequencies.)

Coaxial Magnetron Life. The power that can be produced by a magnetron depends on its size. A larger size means more resonators, which makes it more difficult to separate the various modes of oscillation in a conventional magnetron. The coaxial magnetron, however, with stabilization controlled by the outer cavity, permits stable operation with a larger number of cavities, and thus with greater power. The anode and cathode structures of a coaxial magnetron can also be bigger, which further allows operation at higher power. The larger structures permit more conservative design, with the result that it has longer life and better reliability than conventional magnetrons, as well as more stable operation. The operating life of a coaxial magnetron has been said[26] to be between 5,000 and 10,000 hours, which is a five- to twentyfold increase compared to conventional high-power magnetrons.

Limitations of Magnetrons. When the magnetron was first introduced, it provided a capability not available with the grid-controlled tubes used for early radars. As time passed, the demands for improved radar performance outran the capabilities available from the magnetron. Fortunately, other tube types were invented that overcame the limitations of the magnetron.

Although the magnetron has had important applications in the past, it has also had serious limitations that constrain its usefulness for radar. Its major limitations are its poor stability which limits the ability to detect moving targets in clutter, its relatively modest average power, and its signal cannot be readily modulated for pulse compression. These and others are discussed next.

The use of the doppler frequency shift to detect moving targets in the midst of large clutter echoes requires that the transmitter produce a stable signal with little extraneous noise. Because of their poor stability and noisy transmissions, magnetrons are limited in the amount of MTI Improvement Factor they can achieve to about 30 or perhaps 40 dB. Many radar applications require greater MTI Improvement Factors. Some radars also require the use of pulse-compression waveforms to obtain the resolution of a short pulse with the energy of a long pulse. It is difficult to phase or frequency modulate the waveform of a magnetron, as is needed for pulse compression. Thus, power amplifiers are almost always used for pulse compression applications. Magnetrons are not stable enough to be suitable for very long pulses (e.g., 100 μs), and starting jitter limits their use at very short pulse widths (e.g., 0.1 μs), especially at high power and at the lower frequency bands. Its maximum average power is of the order of several kilowatts, which is less than that required for some military applications.

Since the magnetron is an oscillator with a random starting phase on each pulse, it cannot be used to eliminate second-time-around clutter echoes—as can an amplifier transmitter. Similarly, combining the power outputs of multiple magnetrons has not been attractive. Magnetrons can produce considerable electromagnetic interference across a bandwidth much wider than the signal bandwidth (coaxial magnetrons are somewhat better in this respect). Also, magnetrons do not have precise frequency control nor are they able to perform precise frequency jumping.

In spite of its many unfavorable characteristics, the magnetron is a tube that can be considered for less demanding radar tasks. For a long time, it was the transmitter of choice for use in the civil marine radar, one of the most widely used radars, as briefly discussed next.

Civil Marine Radar Magnetrons.[27] The magnetron has been well-suited for application in civil marine radars used on small pleasure boats or large commercial ships. Its success has been due in part to the radar needing only small transmitter power, and the radar does not require doppler processing to separate moving targets from large fixed clutter echoes. Thus, many of the problems that occur with magnetrons when used in other applications are not found is this application. Also important is that the civil marine radar business is very competitive because of the large worldwide need for such radars. This has resulted in the development of low-cost, highly reliable magnetrons for this important radar application.

These magnetrons generate peak powers from 3 to 75 kW and have relatively low average powers of a few watts to a few tens of watts. An example of a magnetron for a civil marine radar is the MG5241 manufactured by EEY of Chelmsford, England. It is an 18 cavity X-band magnetron that operates at a fixed frequency within the band from 9380 to 9440 MHz with a peak power of 12.5 kW and an efficiency of 43%. It operates with an anode voltage of 5.8 kV and an anode current of 5.0 amps. Typically, its pulse width might be 1.0 μs with a duty cycle of 0.001. The manufacturer claims an expected typical life of over 10,000 hours and guarantees a minimum life of 3,000 hours.

It might also be mentioned that the magnetron has had outstanding success as the power source for the microwave oven. Over the years, it has developed into a very low cost and highly reliable generator of microwave power that is well-suited for this application.

10.4 CROSSED-FIELD AMPLIFIERS[28]

The crossed-field amplifier, or CFA, like the magnetron, has a magnetic field that is orthogonal to the electric field, but it is an amplifier rather than an oscillator.[29] It is similar in appearance to a magnetron except that the RF circuit is interrupted to provide the input and output connections as needed for an amplifier. CFAs might have an efficiency from 40 to 60%, use a lower voltage than linear-beam tubes, are lighter in weight and smaller in size, and have been found from UHF to K band. However, they have relatively low gain and their stability and noise are not as good as found in linear-beam tubes, so their application for MTI radar has been limited. Because of the CFA's low gain, the crossed-field amplifier transmitter needs more than one stage of RF amplification, each with its own power supply, modulator, and controls. All these stages must be stable to achieve good MTI performance.

Because the CFA has relatively low gain, it is sometimes used only in one or two of the highest-power stages of an amplifier chain, where it may offer an advantage in efficiency, operating voltage, size and/or weight compared to other tubes. The output-stage CFA might be preceded by a medium-power TWT that provides most of the gain of the total amplifier chain. (When comparing the CFA with the linear-beam tube, the entire transmitter system needs to be compared and not just the tube itself.) The CFA has also been considered as a means to boost the power output of previously existing radar systems that employed magnetrons. There have been both backward-wave and forward wave CFAs. The backward-wave CFA is also known as the *Amplitron*.

It is possible to pulse some CFAs, which have cold cathodes, to employ what is called *DC operation*, where the transmitter is turned on and off to generate a pulse waveform without the need for a high-power modulator. In dc operation the high voltage is continuously present between anode and cathode, and the current is turned on by applying the RF drive and turned off by pulsing the control electrode. (The control electrode consists of a segment of the cathode structure in the drift region.) To prevent the tube from starting without RF drive, the cathode must be kept cool enough to prevent thermionic emission. The control electrode needs only a short, medium-power pulse, typically one-third of the anode voltage and one-third of the anode peak current. Since the control electrode is insulated and since some energy is dissipated on the control electrode, it can be difficult to cool. This can limit the maximum pulse repetition frequency that can be used. In spite of it requiring no modulator, dc operation has seldom been used because of its many limitations.[30]

Crossed-field amplifiers have been used in radars in the past, but they have significant disadvantages, as was described in the second edition of this *Radar Handbook,* that make it less likely they will be widely used in the future.

10.5 GYROTRONS[31,32,33]

It has been noted previously that the power handling capability of the microwave power tubes discussed thus far in this chapter decreases as the frequency is increased. This results because the resonant structures of the slow-wave microwave circuitry of these tubes become smaller with increasing frequency. The smaller the device the more difficult it is to dissipate the heat that is generated. Thus, the power output decreases approximately inversely as the square of the radar frequency.

The gyrotron RF power generator, however, does not have this limitation because it does not employ slow-wave resonant microwave structures. These devices employ a fast-wave structure, such as a smooth circular tube,* one where the phase velocity of the electromagnetic wave is greater than the velocity of light. (With a slow-wave device, the phase velocity is less than the velocity of light.) The diameter of the gyrotron circuit can be many wavelengths, and the electron beam need not be placed close to delicate RF structures. Because a fast-wave structure rather than a slow-wave structure is used, they do not have the size limitations of other microwave power sources as the frequency is increased. Thus, they can generate greater power at the higher frequencies than can other RF power sources, something especially attractive at millimeter wavelengths.

* *Tube* is used here as being a "hollow elongated cylinder."

The magnetic field in a gyrotron serves a different function from the magnetic field in a slow-wave device. In the slow-wave device, the magnetic field keeps the electronic beam collimated. In the fast-wave device, however, the magnetic field determines the frequency; but the frequency of a conventional slow-wave device is determined by the circuit dimensions. An electron in an applied axial magnetic field B_o will rotate at what is called the *electron cyclotron frequency,* which is given by $\omega_c = eB_o/m\gamma$, where e = electron charge, m = electron rest mass, and γ is the relativistic factor, which is $[1 + (e/mc^2)V_o]$, where c = velocity of light and V_o = beam voltage. The beam voltage and the corresponding electron velocity in a gyrotron are high enough to cause relativistic effects. The electrons follow helical paths around the magnetic field lines when in the presence of an electromagnetic wave with a transverse component of electric field. Electrons that lose energy to the electromagnetic wave become lighter and accumulate phase lead and then catch up with the electrons that gain energy and become heavier and accumulate phase lag. Thus, electrons become phase-bunched in their cyclotron orbits as a result of the relativistic mass change of the electrons. The gyrotron bunching operation also can be obtained at harmonics of the cyclotron frequency, but there can be problems with higher circuit losses and competition with modes operating at lower harmonics so that most high-power gyrotrons operate at the fundamental frequency or its second harmonic.[33]

Because the frequency of a gyrotron is determined by the magnetic field and not by the size of the fast-wave structure, the structure can be large, and it is then possible to generate quite high power at millimeter-wave frequencies. The large magnetic fields needed for millimeter-wave gyrotrons often have to be generated by superconducting magnets.

The gyrotron with a single cavity operates as an oscillator. It is sometimes called a *gyro-oscillator* to differentiate it from a *gyro-amplifier* that utilizes several resonant cavities or a traveling wave circuit to operate as an amplifier. The gyro-amplifier that employs several resonant cavities is called a *gyroklystron*, and when a traveling-wave circuit is used, it is called a *gyro-traveling-wave-tube*, or, more commonly, *gyro-TWT*. There is also a *gyrotwystron* with the resonant output cavity replaced by a TWT circuit so as to achieve greater bandwidth than can be obtained with a resonant cavity. Although gyro-oscillators have been capable of greater power than gyro-amplifiers, the gyro-amplifier has usually been preferred for radar applications for the same reason the amplifier has been preferred at microwave frequencies, especially when doppler processing is important.

An example of a high power gyroklystron for radar applications is the VGB-8194[34,35] that was used in the experimental W-band radar known as *Warloc* at the U.S. Naval Research Laboratory. It had five cavities, operated over a 700 MHz bandwidth centered at 94.2 GHz with an average power of 10.2 kW, a peak power of 102 kW, 10% duty cycle, and 31% efficiency with a 55 kV, 6 A electron beam. It used a superconducting magnet with a closed-cycle cooling system (so that no liquid cryogens were required). It provided a magnetic field of 36.6 kG. The five-cavity gyroklystron could achieve a bandwidth of 1050 MHz at an average power of 4 kW. The radar used a 6 ft diameter antenna that provided a beamwidth of 0.1°. This W-band Warloc was installed in a van and was used for various experiments. The Warloc radar was about three orders of magnitude more capable than most previous radars used for millimeter-wave radar. This experimental Warloc radar was employed to demonstrate at W band the ISAR imaging of moving targets, cloud structure, low-angle operation, and unusual atmospheric research including what have been called “air spikes.”

The above discussion was mainly about the gyrotron amplifier. The gyrotron has also been operated as an oscillator to produce very high power, but the oscillator version has not been as popular as the amplifier for radar applications. This might be due to the amplifier being better able to produce the desired radar waveforms.

10.6 TRANSMITTER SPECTRUM CONTROL

The increasing demand for electromagnetic spectrum for both civilian and military applications has accentuated the need to control the spectrum of radar transmitters to avoid interference with users of the electromagnetic spectrum operating at other frequencies. The aspects of concern here are those that affect RF tube selection and achievement of minimum spectrum occupancy by the transmitters discussed previously in this chapter.

Reduction of Spurious Outputs. RF tube spurious outputs may be grouped into three kinds: harmonics, adjacent-band, and in-band.

All RF tubes produce some harmonic output. In general, little can be done in tube design to reduce harmonic outputs, but it is feasible to filter out harmonics (30 to 60 dB reduction) with high-power filters.

Adjacent-band spurious output can also occur in some cathode-pulsed TWTs and CFAs. Adjacent-band spurious output is affected by tube and modulator selection, but it can be filtered by a high-power microwave filter if necessary.

All RF tubes produce some in-band background noise level. In a 1-MHz bandwidth, this noise might be 50 to 60 dB down in conventional CFAs, 70 to 80 dB down in the low-noise-high-gain CFA, and 90 dB down or better in linear-beam tubes. In-band spurious cannot normally be improved with filters because it occurs within the same frequency range as the desired signal spectrum. Attempts to use noise degeneration to reduce the inherent RF tube noise levels are subject to limitations. In-band spurious signals can also result from power supply and modulator instabilities.

Reduction of Spectrum Amplitude Exceeding (sin *x*)/*x*. The spectrum of a perfectly rectangular pulse has the familiar (sin *x*)/*x* form, where x is $\pi(f_o - f)\tau$, f_o is the radar carrier frequency, and τ is the pulse width. If $1/\tau$ is called the nominal bandwidth of the signal, the envelope of the spectrum peaks falls off at the rate of 6 dB per octave of bandwidth, and this reduction will continue until the envelope reaches the inherent noise output level of the transmitter. This rate of spectrum falloff is too slow to meet most system requirements. Nevertheless, without special care the actual spectrum envelope might be even worse than this, depending on tube characteristics, as a result of phase modulation during the finite rise and fall of practical modulator and RF drive pulse shapes. In these cases, either the leading and trailing edges must be appropriately tailored, or else (in linear-beam tubes) the RF drive may be withheld during the rise and fall time. Although this may slightly reduce apparent efficiency, it should be noted that energy outside the approximately $1/\tau$ generated during rise and fall with the RF drive present is not utilized by the receiver anyway.

Improvement by Means of Shaped Pulses. Since the energy in the spectrum beyond plus or minus $1/\tau$ from f_o is not used by the receiver, it is desirable for

electromagnetic compatibility purposes to avoid transmitting energy beyond those limits. This objective may be approached by using a pulse shape different from the convenient and conventional rectangular pulse.[36] Highly shaped pulses have not been used often in radar systems because of the loss of efficiency that results. (These limitations, however, do not apply when using the grid-controlled tubes known as the *Constant Efficiency Amplifier* or the *Inductive-Output Tube,* as discussed later in Section 10.7.)

Another approach to spectrum improvement is to shape the rise and fall of a rectangular pulse.[37] This attenuates the spectrum of frequencies far from f_o, while the flat-topped portion of the pulse retains the high transmitter efficiency for most of the pulse duration. Since a rectangular pulse has the best transmitter efficiency but has high spectral energy at frequencies far from the center frequency, whereas a highly shaped pulse has less far-out spectral energy but poor transmitter efficiency, the fraction of the pulse length to be used for the shaped rise and fall is a crucial decision.

Although the improvement attainable in practice is limited by phase modulation in the transmitter during the rise and fall,[38,39] significant improvements can be obtained. In a linear-beam-tube transmitter with properly shaped RF drive, for example, the spectrum width at 60 dB down can usually be narrowed by about an order of magnitude at a cost of about 1 dB in transmitter efficiency.

Amplifier chain radar systems, whether tube or solid state, often use some shaping of the edges of the transmitted RF pulse to reduce the RF spectrum width. This can be done by simply slowing the rise and fall times of the exciter signal to the transmitter, and this approach generally has been adequate to satisfy Military Standards and related system requirements.

Spectral Noise in Doppler Radars. The doppler frequency shift is widely used to detect moving target echo signals in the presence of large clutter echoes. However, if the radar transmitter generates noise or its pulse waveform has significant spectral energy at the doppler frequencies expected from moving targets, this unwanted noise or spectra will degrade the detection of desired targets. The transmitter noise is reflected back from the clutter and enters the receiver and is sometimes called "transmitted clutter."[40] Some types of microwave tubes are more of a problem than others. Extraneous noise in a microwave amplifier tube at the frequencies expected from doppler-shifted target echoes can be produced by ion oscillations. According to A. A. Acker,[40] "these are periodic instabilities that can occur in the electron beam at video frequencies, resulting in signals other than the carrier frequency, causing severe problems to doppler radar performance." Acker also states that ion oscillations require a finite time to develop so that if a tube is operated with less than a 10 μs pulse width, ion noise is usually not of concern.

Advances in digital technology have allowed a method to reduce intra-pulse transmitter noise and power supply instabilities that affect radar performance by the presence of strong clutter echoes. The technique,[41] known as *Transmitter Noise Compensation (TNC),* captures and processes an accurate replica of each transmitted pulse. By means of pulse-to-pulse comparisons, the measured transmitter errors are used to derive a digital filter that compensates for the transmitter noise that arrives in the receiver digital signal processor. TNC compensates for intra-pulse transmitter noise as well as power supply instabilities. Although TNC works only in a single unambiguous range interval, it is said that it should be able to operate with some medium PRF radars if significant clutter is not likely to extend over more than one PRF interval.

An experimental implementation using data collected on an operational radar with a CFA transmitter showed that "the TNC technique can improve radar detection of targets in clutter by 15 dB or more."

10.7 GRID-CONTROLLED TUBES

The grid-controlled tube is a modern version of the classical triode or tetrode vacuum tube that dates back to the early years of the 20th century. These devices employ a cathode to generate electrons, one (for a triode) or two (for a tetrode) control grids, and an anode to collect the electrons. A small voltage applied to the control grid acts to control the number of electrons traveling from cathode to anode. The process by which the electron density of the electron stream is modulated by the signal on the control grid to produce amplification is called *density modulation*. In the latter half of the 20th century, grid-controlled tubes were successfully employed in such important radar applications as HF over-the-horizon radar, VHF and UHF aircraft surveillance radars, and satellite surveillance radars. Grid-controlled tubes are capable of high power, wide bandwidth, good efficiency, and inherent long life, but they are of low or moderate gain. Their chief limitation is that they are not capable of operating at higher frequencies, but they are capable up to frequencies approaching 1 GHz. Grid-controlled tubes have been operated at UHF and higher frequencies by using microwave techniques in their construction, as was done in the Coaxitron. They can also be made to operate with constant efficiency when shaping of the pulse amplitude is used to reduce interference caused by its far-out spectrum, something not as practical to do with other types of microwave tubes.

Coaxitron. The performance of conventional grid-control tubes at the higher frequencies is limited by the time it takes for the electrons to transit from the cathode to the anode of the tube. This transit time must be small compared to the period of the RF signal to be amplified. To minimize the undesired transit-time effects, the complete RF input and output circuit and the electrical interaction system can be placed within the vacuum envelope. Such a grid-controlled tube is called a *Coaxitron*.[42] In one embodiment of the Coaxitron, the electron-interaction structure consisted of a cylindrical array of 48 essentially independent grounded-grid unit triodes.

The so-called A15193F Coaxitron, described in Vingst et al.,[42] operated over a frequency range from 406 to 450 MHz with a peak power of 1.25 MW, a pulse width of 13 μs, a duty cycle of 0.0039, an average power of just under 5 kW, and a plate efficiency of 47%. The Coaxitron has been successfully employed for UHF radars, including airborne.

Constant Efficiency Amplifier (CEA). It has been said[43] that "a constant-efficiency amplifier has been a goal for transmitter engineers ever since Lee DeForest and Ambrose Fleming invented the first electronic amplifiers." This goal seems to have been achieved by the CEA grid-controlled tube.

It is customary to think of the shape of a conventional radar pulse as being rectangular. It is seldom, however, perfectly rectangular with very short rise and fall times because such a waveform has a very large bandwidth, as one can observe from its Fourier transform. Even if a large bandwidth were available to support a

rectangular pulse, a large bandwidth would likely cause interference to other radars and other electromagnetic systems. For this reason, government frequency allocation agencies usually require that the frequency spectrum from a radar not contain large energy at other frequencies. This is becoming more important as the occupancy of the electromagnetic spectrum is increasingly crowded with transmitters. The classical way to reduce the far-out spectrum from a radar transmitter is to shape, or taper, its waveform, such as by using a trapezoidal shape, a gaussian-like pulse shape, a truncated gaussian, perhaps a cosine on a pedestal, or other type of nonrectangular shape. The problem with using conventional transmitters such as discussed in this chapter is that when using a shaped waveform some loss in efficiency results. Thus, it is seldom that a radar designer would want to use a highly shaped pulse waveform in order to reduce the spectrum radiated outside of the radar's normal operating signal bandwidth. The CEA, however, is an RF power source that does not have its efficiency decreased when a nonrectangular, or shaped, waveform is used. The CEA has been widely used for commercial TV transmitters that have highly modulated, nonconstant amplitude waveforms.

The CEA is based on a grid-controlled tube known as an *Inductive Output Tube,* or IOT. (The CEA is similar to something called a Klystrode[44] except that the CEA employs the IOT with a multistage depressed collector similar to that used in klystrons and TWTs.[37]) In the IOT, the wire grid of a grid-controlled tube is replaced with an aperture that does not intercept the electrons, and it has a coaxial magnetic field that confines the electron stream as in a klystron or a TWT. Although an RF cavity is used in an IOT, the beam is density modulated, or bunched, with a grid similar to how it is modulated in a triode or tetrode grid-controlled tube. This makes it smaller and lighter than a comparable klystron. The density modulated electrons thus form bunches, and RF energy is extracted by passing the beam through the resonant cavity. The CEA has been widely used for UHF-TV transmitters, a highly competitive industry concerned with cost. It has been said that in UHF-TV, the use of a CEA reduces the required prime power by one-half compared to a conventional tube transmitter[45] and by one-third of the prime power of a silicon-carbide solid-state transmitter.[46] The CEA achieves these efficiencies since there is no loss of efficiency when using shaped pulse waveforms, as there is with other microwave tubes. This is an important reason for its use for UHF-TV with its time-varying-amplitude waveform. It also should be an advantage for radar applications that require shaped waveforms at frequencies up to 1000 MHz, when it is required to reduce the far-out spectral energy.

A CEA, such as manufactured by L-3 Communications, can operate within the UHF-TV band from 470 to 806 MHz with an 8 MHz bandwidth (the spectral width of a TV channel), 130 kW peak power at 60% efficiency, and an average power of 6 kW or greater. The single input cavity can be made to tune over the entire band with low VSWR.

Thus, the constant efficiency amplifier is a grid-controlled tube, operating as class-AB, that consists of an inductive-output tube with a multistage depressed collector. It is a linear amplifier whose prime power can be proportional to the output power, providing constant efficiency over a wide range of output powers. The CEA seems to be the preferred tube for UHF-TV transmitters, rather than other types of grid-controlled tubes, solid-state, or klystrons. It would seem that the CEA ought to be of interest for radar applications at frequencies as high as 1000 MHz when highly shaped pulses need to be used.

Application of Grid-Controlled Tubes. There has been important application of the grid-controlled tube for radars in the past at HF, VHF, and UHF. It is still of value for applications in these frequency regions and should be considered as a candidate when designing a radar to operate at the lower frequencies. The constant efficiency amplifier should be of interest because of its higher efficiency when shaped waveforms are needed for control of the radiated spectrum.

10.8 MODULATORS

This section briefly reviews the *modulator,* sometimes called the *pulser,* which is the device that turns the transmitter tube on and off to generate the desired pulse waveform. Those desiring more information might examine Chapter 9 in L. Sivan's book on transmitters,[47] the chapter by T. A. Weil in the second edition of *Radar Handbook,*[48] the Conference Records of the International Power Modulator Symposia, and the Conference Proceedings of the IEEE Pulse Power Conferences.

The type of tube determines, to some extent, the type of modulator. A modulator basically consists of an energy storage device, which may be a capacitance or a pulse forming network, and a switch for triggering the dc pulse. The switch in the past might have been a vacuum tube, thyratron, ignitron, silicon-controlled rectifier (SCR), reverse-switching rectifier, spark gap, or magnetic. However, the solid-state switch seems to be the switching mechanism that should be considered initially when designing a transmitter modulator. Modulators can be classed as either low power or high power, depending on how the tube is modulated.

If the tube has a grid, a small and relatively inexpensive type of modulator can be used, but grids usually are not feasible with high power tubes. A widely used switching element for low-power modulators is the MOSFET transistor.[47]

Low-power modulators can be used with tubes that employ a modulating anode, as does the linear-beam amplifier. The modulating anode in a linear-beam tube is part of the electron gun and is separated from the body of the tube. The voltage of the modulating anode is varied over a large range in order to vary the electron-beam current, but the power needed to drive the modulating anode is small because the current intercepted by the modulating anode is very small.

Very high-power tubes cannot use a modulating anode because the control electrode might not be able to handle the power. When this occurs, a high-power modulator, called a *cathode pulser* might have to be used to switch the tube on and off. Cathode pulsers must switch the full beam voltage and current simultaneously, which involves high instantaneous powers. They must control the full beam power of the RF tube, either directly or through a coupling circuit. The energy storage device might be made up of capacitors, inductors, or some combination of the two as in pulse-forming networks. The energy in the energy storage device is discharged by a very capable switch.

The *line-type modulator* uses a delay line or a pulse-forming network (PFN) as the energy storage element. A switch initiates the discharge of the energy stored in the PFN. The shape and duration of the pulse are determined by the passive elements of the PFN. The switch has no control over the pulse shape, other than to initiate it. The pulse ends when the PFN has discharged sufficiently. A disadvantage of this action is that the trailing edge of the pulse is usually not sharp since it depends on the discharge characteristics of the PFN. It has been widely used in the past for magnetron pulsing.

In an *active-switch modulator,* the switch has to be turned off as well as turned on. Originally, the switch was a vacuum tube and the modulator was called a hard-tube modulator to distinguish it from the gas-tube switch often used in a line-type modulator. Since other than vacuum tubes can be used as the switch in an active-switch modulator, the "hard-tube" designation (meaning a vacuum tube) might not always apply. Unlike the line-type modulator, the switch in the active-switch modulator controls both the beginning and end of the pulse. Since the energy storage device is a capacitor, the pulse can droop, something that can be prevented by extracting only a small fraction of the stored energy from the capacitor. This requires a large capacitance, which may be obtained with a collection of capacitors called a *capacitor bank*. The active-switch modulator permits greater flexibility and precision than the line-type modulator. It can provide excellent pulse shape, varying pulse durations, and pulse repetition frequencies, including mixed pulse lengths and bursts of pulses with close pulse spacings.

Microwave tubes and their high-voltage switches can sometimes produce an unwanted arc discharge that effectively places a short circuit across the power supply and/or modulator delivering power to the tube. Since 50 J of energy can usually cause damage to an RF tube (or the switching device) and since the capacitor bank for an active-switch modulator must often store far more than 50 J (to prevent excessive droop), some means must be provided to divert the stored energy when an arc discharge occurs. Such a device is the *crowbar,* so called because it is equivalent to placing a heavy conductor (like a crowbar) directly across the power supply to divert the energy and thus prevent the energy from discharging through the tube and causing serious damage. A crowbar is, therefore, needed for a high-power active-switch modulator because of the large amount of energy stored in its capacitor bank. On the other hand, crowbars are not usually needed with line-type modulators, which store less energy in their pulse-forming network.

Certain crossed-field amplifiers can be pulsed by means of a control electrode, located in the tube's drift region, without a separate full-power pulse modulator.[49] This is called *dc operation*. Even though dc operation avoids a high power modulator, it has seldom been used because it requires a much larger capacitor bank to limit droop, an arc in the tube requires a crowbar that interrupts operation for a few seconds instead of only for a single pulse, and the adjacent radars might inject enough RF power into the radar antenna and back to the transmitter to turn on a dc operated CFA at the wrong times. There has been at least one example in the past where a major radar system originally designed with CFAs based on dc operation had to have its dc operated CFAs replaced during the middle of its development with CFAs that used conventional pulse modulators.

At the beginning of the 21th century, the *solid-state modulator* was developed and began to be used for radar transmitters as either cathode pulsed or modulating anode pulsed modulators, as well as grid pulsed. Solid-state modulators offer improved transmitter performance by allowing a wide variation in parameters (pulse width, pulse repetition frequency, pulse agility, and pulse-to-pulse consistency). They also provide cost savings that result from the inherent reliability of these switch modules compared to conventional switch tubes, and the elimination of numerous auxiliary components needed for the operation of switch tubes.[50,51] Lower operating costs and smaller cooling requirements occur because of their higher conversion efficiency. They are also said to have higher reliability with longer component life. The ability of the solid-state switch to open quickly (less than one microsecond) when a fault is detected eliminates the need for a crowbar. The energy-storage device does not discharge during an arc, so when the fault has cleared, the transmitter can resume operation in microseconds.

A solid-state cathode modulator can provide pulse widths varying from 50 ns to "dc" on a pulse-to-pulse basis and can support pulse repetition frequencies up to 400 kHz.[52] The high voltage solid-state switches are built from modules that might contain from 4 to 20 individual transistors connected in series to provide the required transmitter cathode voltage. Rise times can be as low as 30 ns.

10.9 WHICH RF POWER SOURCE TO USE?

There is no good, simple answer to this question, but in this section we shall attempt to discuss some of the various issues that might be involved.

This chapter has briefly described the various vacuum tubes that have been used or considered for radar applications, and the next chapter discusses the solid-state transmitter, which has also been widely used in radar. A question that naturally arises is which RF power source should be used for some particular radar application. Radar system design usually involves making choices among the various possibilities available. When trying to determine which RF power source to use, a choice can be made by doing a separate system design with each RF power source that seems promising. The decision as to which to use can be based on how well each system design performs the desired task according to some pre-established criteria. This, unfortunately, is seldom done. It is suspected that sometimes the decision as to which RF power source to use is determined by what the radar system designer thinks the buyer (or customer) of the radar desires. Sometimes the buyer will actually specify the type of transmitter to be delivered. Manufacturing a product based on what the buyer thinks he or she wants may be a good marketing strategy for many products, but in something like a radar, it might be better for a customer to clearly specify what performance is wanted and then leave the decision as to what RF power source should be used to the radar system designer. It is usually better if the radar design can be determined by the radar system designer and not by the manufacturer's marketing department. The goals of the radar designer and the marketing department are not always the same. It is appreciated, however, that sometimes the marketing manager's opinion has to prevail if the company is to remain in business.

A number of different RF power sources have been considered that, at one time or other, have been employed in radar. Not all might be popular or desired at some particular time but most should be considered, even if briefly, by the radar system designer when trying to determine a new radar system design or an upgrade of some existing system. Opinions about the utility of the various vacuum tube transmitters mentioned here will be briefly given next, with the suggestion to the reader to keep in mind that circumstances can change and these opinions can change as well. These opinions are not "written in stone," and are not likely to be universally agreed to by all those who work in radar. But that is the nature of any engineering endeavor.

Brief Opinions About the Utility of Various Radar Vacuum Tubes. The types of RF power sources are mentioned below in no particular order.

Grid-Controlled Tube. Although some might think these should have disappeared along with the old radio vacuum tube, there have been many HF, VHF, and UHF radars that successfully operated with grid-control tubes. Often with such radars, it would cost more to replace them with solid-state transmitters, and there might be little gained by doing so. The grid-controlled tube known as the constant efficiency amplifier (CEA)

and its predecessors (the IOT and the Klystrode) are the only RF power sources that can operate efficiently when amplitude shaped waveforms are desired for minimizing out-of-band interference to other radars. Thus, the CEA probably should be a candidate when considering any new UHF radar system, as well as radars at lower frequencies, especially if mutual interference is a potential problem.

Magnetron. It was mentioned that the magnetron was what made microwave radar possible in the 1940s. The magnetron is still a valid candidate for small, non-doppler radars such as civil marine radars, although such radars also have been manufactured with solid-state transmitters. It is not likely that magnetrons will be used in high performance radars, especially those that require average powers more than one or two kilowatts or where MTI improvement factors have to be greater than 30 to 40 dB. For example, during the procurement of the Nexrad doppler weather radar in the mid-1980s, the magnetron was considered, but it could not meet the clutter cancellation specifications, which is why Nexrad employs a klystron. In the past, some long-range Air Traffic Control Air Route Surveillance Radars used a magnetron, but the klystron seems to be the preferred choice for this application.

Crossed-Field Amplifier. Although these tubes were employed for some major radar applications because they have good efficiency, require relatively low voltage, and have wide bandwidth (about 10%), they are less likely to be used because they are noisy (which affects their doppler-processing performance), they are of relatively low gain (which requires the transmitter to have multiple stages), and because the klystron is usually a better overall choice.

Klystron. The original klystrons employed resonant cavities that restricted their bandwidth. The bandwidth of a klystron, however, increases as its power increases. Resonant cavities were later replaced by wider-band circuits, which were related to the type of circuits used in TWTs. Such klystrons are known as the Clustered Cavity Klystron, Extended Interaction Klystron, and the Twystron. When considering a transmitter for high-performance radar, these variants of the klystron are likely to be highly favored for many applications. The klystron has good stability and low noise so as to enable larger MTI improvement factors to be obtained when using the doppler shift to detect moving targets in clutter. At high power, high voltages have to be used and protection from X-rays generated by the high voltage must be employed. However, the MBK (multiple-beam klystron) version of the klystron can be used to achieve high power with lower voltage.

Traveling Wave Tube. As mentioned, the TWT and the klystron can have comparable bandwidths when the tube produces high power. The performance of a TWT is similar to that of a wideband klystron, except that it might not be as stable as the klystron and have slightly less gain. The microwave power module (MPM), which is a combination of helical TWT and solid-state device, has not had significant application in radar.

Gyrotrons. If very high power is needed at millimeter-wave frequencies, the gyrotron amplifier or oscillator is the only RF power source available. For low-power radar applications at millimeter waves, the EIK can be used.

Solid-State Amplifiers for Radar Transmitters. Solid-state transmitters and vacuum tube transmitters have significant differences, yet they are both employed

in radar. Some of these differences are mentioned in Section 11.1 of Chapter 11 "Solid-State Transmitters." Briefly, proponents of solid state might say that they do not need a hot cathode as does a vacuum tube, do not require high voltages or magnets, do not produce X-ray radiations as some vacuum tubes do, have "graceful degradation," and that maintainability is its key asset. On the other hand, proponents of vacuum tubes might say that radars using solid-state transmitters have limited peak power and thus need to operate with a long pulse and a high duty cycle, which require the use of pulse compression. The long pulse can mask or eclipse target echoes at short range so that an additional short pulse transmission is needed to unmask the eclipsed echoes at short ranges. When Sensitivity Time Control (which has a varying receiver gain with range) is used with a long pulse and pulse compression, distortion can result in the compressed pulse. It has also been said that solid-state transmitters are often less efficient, they might be heavier, and their cost might be greater than an equivalent vacuum tube radar system. The above have all been said at one time or another, but there is not universal agreement about the importance of these characteristics for all radar applications.

The radar engineer should not simply compare the particular differences between a solid-state transmitter and a vacuum tube transmitter when determining what type of RF power source to use in any particular application. The choice between the two should be made by comparing a radar system designed to effectively use solid state and a radar system designed to effectively use vacuum tubes. Assuming the solid state and vacuum tube radars are designed to provide identical performance for the desired application, then the choice should be based on comparing cost, size, weight, reliability, maintainability, and any other system requirement that is important for decision making. Unfortunately, this is not always done. Buyers of radars should be encouraged not to insist that the radar designer use some particular technology because it is considered the fashionable thing to do at the time. They might not always be getting the best radar for the particular application.

There have been at least three ways to apply solid-state transmitters to high-performance radars: (1) as a replacement for a vacuum tube transmitter in an already existing radar; (2) as the transmitter for a new radar design; and (3) as an active aperture phased array radar.

An example of replacing an existing vacuum tube transmitter with a solid-state transmitter is the U.S. Navy's AN/SPS-40, a relatively modest capability UHF radar for air surveillance.[53] This radar was chosen for having its vacuum tube transmitter replaced with a solid-state transmitter using transistor amplifiers because it already used a long pulse waveform with a high duty cycle and pulse compression, which is what solid-state radars usually require. The solid-state transmitter was put into production and installed in existing radars. It did the job it was supposed to do, but it is not obvious that the solid-state transmitter had a net advantage over the tube transmitter. The solid-state transmitter was supposed to occupy the same floor space as the vacuum tube transmitter, but it occupied about the same floor space as the entire AN/SPS-40 radar, which used a vacuum tube. Also the solid-state transmitter cost more than the tube version. One highly significant advantage of the solid-state version of the SPS-40, however, was that it could include spare solid-state modules as part of the transmitter itself, so that time to repair was reduced.

An example of the second approach to achieving a solid-state radar transmitter is the ASR-12 airport surveillance radar. In the mid-1980s, the ASR-9 air surveillance radar at S band was developed by Northrop Grumman (then Westinghouse) for use at major airports to control local air traffic.[54] It was an excellent radar that used a

well-tested klystron amplifier vacuum tube, and it was installed throughout the United States. (The same tube is used in the Nexrad doppler weather radar.) However, in the late 1990s solid-state technology had advanced sufficiently so that Northrop Grumman developed the ASR-12, also at S band, using a solid-state transmitter. Its overall radar performance was similar to the ASR-9, but it was not just a replacement of the transmitter but a new design to use solid-state transmitters effectively. It also took important advantage of advances in digital receivers and digital processing that occurred since the development of the ASR-9 to significantly improve what could be accomplished with this radar.[55] As mentioned, solid-state transmitters require the use of long pulses. The ASR-12 employed a 55 μs pulse duration at a peak power of 21 kW. This means that targets out to a range of about 5 nmi would be masked, or eclipsed, by the long pulse and might not be detected. To detect targets at ranges masked by the long pulse, a second (short) pulse 1 μs in duration, and at a different frequency from the long pulse, was transmitted almost immediately after the long pulse. It detects targets within the range from 0.5 nmi or less to a range of about 5.5 nmi. The long pulse employs nonlinear FM pulse compression with a 55:1 pulse compression ratio to achieve a range resolution of less than 1/8 nmi as required for an air-traffic control radar. Typical time-sidelobes with the nonlinear FM waveform were 58 dB below the peak response. It might be noted that Cole et al.[55] state that in order "to ensure continued availability of the power transistors required to produce the power amplifier panels, Northrop Grumman has developed an in-house manufacturing capability for high-power S-band transistors."

The third approach to employing solid-state transmitters is the active aperture phased array radar. At each element of a phased array radar antenna is a solid-state module, known as a *T/R module,* that contains a transmitter, receiver, and duplexer. The vacuum tube is not usually competitive for this application. In Chapter 5, "Multifunctional Radar Systems for Fighter Aircraft," the active aperture radar is called an *Active Electronically Scanned Antenna* (AESA). The subsection entitled "Active Electronic Scanned Array (AESA)" (in Section 5.1) describes quite well the military airborne application of solid-state radar, and enumerates its advantages and why it is important. There it is stated that "one of the principal advantages of an AESA is the ability to manage both power and spatial coverage on a short-term basis (10s of ms)." It is also said that "bandwidth of several GHz on transmit" is required, and this is within the capability of solid-state transmitters. The reader is referred to Section 5.1, Chapter 11, and Section 13.10 for further information about this important application of solid-state.

Although any of the RF power sources mentioned here could be used in future radar systems, it seems likely that the linear-beam amplifier, particularly one of the variants of the klystron, might be the first RF power source to consider for a high performance microwave radar that employs a mechanically steered antenna or a conventional phased array radar that does not employ the active aperture. For active aperture phased array radars, it is likely that the solid-state transistor amplifier will be the choice.

REFERENCES

1. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill 2001, p. 88.
2. A. S. Gilmour, Jr., *Microwave Tubes*, Norwood, MA: Artech House, 1986.
3. W. J. Dodds, T. Moreno, and W. J. McBride, Jr., "Methods for increasing the bandwidth of high power microwave amplifiers," *IRE WESCON Conv. Rec*. 1, pt. 3, 1957, pp. 101–110.

4. R. J. Barker et al., *Modern Microwave and Millimeter-Wave Power Electronics*, New York: IEEE Press and Willey Interscience, 2005, p. 108.
5. A. S. Gilmour, Jr., *Principles of Traveling Wave Tubes*, Boston, MA: Artech House, 1994, Sec. 18.4.
6. R. S. Symons, "Tubes: Still vital after all theses years," *IEEE Spectrum*, vol. 35, pp. 52–63, April 1998.
7. R. M. Phillips and D. W. Sprehn, "High-power klystrons for the next linear collider," *Proc. IEEE*, vol. 87, pp. 738–751, May 1999.
8. R. H. Abrams, B. Levush, A. A. Mondelli, and R. K. Parker, "Vacuum electronics for the 21st century," *IEEE Microwave Magazine*, pp. 61–72, September 2001.
9. G. S. Nusinovich, B. Levush, and D. Abe: "A review of the development of multiple-beam klystrons and TWTs," Naval Research Laboratory, Washington, DC, MR/6840-03-8673, March 17, 2003.
10. R. H. Abrams, B. Levush, A. A. Mondelli, and R. K. Parker, "Vacuum electronics for the 21st century," *IEEE Microwave Magazine*, pp. 61–72, September 2001.
11. A. N. Korolyov, E. A. Gelvich, Y. V. Zhary, A. D. Zakurdayev, and V. I. Poognin, "Multiple-beam klystron amplifiers: performance parameters and development trends," *IEEE Trans*., vol. PS-32, pp. 1109–1118, June 2004.
12. R. J. Barker et al., *Modern Microwave and Millimeter-Wave Power Electronics*, New York: IEEE Press and Willey Interscience, 2005, Sec. 3.5.3.
13. W. H. Yocom, "High power traveling wave tubes: Their characteristics and some applications," *Microwave J*., vol. 8, pp. 73–78, July 1965.
14. A. S. Gilmour, Jr., *Principles of Traveling Wave Tubes*, Boston, MA: Artech House, 1994, Sec. 18.4.
15. H. G. Kosmahl, "Modern multistage depressed collectors—A Review," *Proc. IEEE*, vol. 70, pp. 1325–1334, November 1982.
16. A. S. Gilmour, Jr., *Microwave Tubes*, Norwood, MA: Artech House, 1986, Sec. 12.2.
17. M. J. Smith and G. Phillips, *Power Klystrons Today*, New York: John Wiley, 1995, Sec. 7.2.3.
18. A. E. Staprans, W. McCune, and J. A. Ruetz, "High-power linear-beam tubes," *Proc. IEEE*, vol. 61, pp. 299–330, March 1973.
19. A. S. Gilmour, Jr., *Microwave Tubes*, Norwood, MA: Artech House, 1986, Sec. 11.3.
20. A. Roitman, D. Berry, and B. Steer, "State-of-the-art W-band extended interaction klystron for the CloudSat program," *IEEE Trans*., vol. ED-52, pp. 895–898, May 2005.
21. R. S. Symons and J. R. M. Vaughan, "The linear theory of the clustered cavity klystron," *IEEE Trans.*, vol. PS-22, pp. 713–718, October 1994.
22. R. S. Symons, "Tubes: Still vital after all these years," *IEEE Spectrum*, vol. 35, pp. 52–63, April 1998.
23. R. H. Abrams, Jr., "The microwave power module: A 'supercomponent" for radar transmitters," *Record of the 1994 IEEE National Radar Conf*., Atlanta, GA, pp. 1–6.
24. C. R. Smith, C. M. Armstrong, and J. Duthie, "The microwave power module: A versatile building block for high-power transmitters," *Proc. IEEE*, vol. 87, pp. 717–737, May 1999.
25. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2001, Sec. 10.4.
26. N. Butler, "The microwave tube reliability problem," *Microwave J*., vol. 16, pp. 41–42, March 1973.
27. P. D. L. Williams, *Civil marine radar*, London: Institution of Electrical Engineers, 1999, Sec. 10.3.
28. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill Companies, 2001, Sec. 10.5.
29. A. S. Gilmour, Jr., *Microwave Tubes*, Norwood, MA: Artech House, 1986, Sec. 13.3.
30. L. L. Clampitt, "S-Band amplifier chain," Raytheon Company, Waltham. MA, presented at NATO Conf. Microwave Techniques, Paris, March 5, 1962.
31. V. L. Granatstein and I. Alexoff, *High Power Microwave Sources*, Boston: Artech House, 1987.
32. A. S. Gilmour, Jr., *Microwave Tubes*, Norwood, MA: Artech House, 1986, Chap. 14.

33. K. L. Felch et al., "Characteristics and applications of fast-wave gyrodevices," *Proc. IEEE*, vol. 87, pp. 752–781, May 1999.
34. M. Blank et al., "Development and demonstration of high-average power W-band gyro-amplifiers for radar applications," *IEEE Trans*. vol. PS-30, pp. 865–875, June 2002.
35. G. J. Linde et al. (private communication), "Warloc: A high-power coherent 94 GHz radar."
36. J. P. Murray, "Electromagnetic compatibility," Chap. 29 in *Radar Handbook*, 1st Ed., 1970, pp. 29.18 to 29.23.
37. T. A. Weil, "Efficient spectrum control for pulsed radar transmitters," Chap. 27 in *Radar Technology*, E. Brookner (ed.), Norwood, MA: Artech House, 1977.
38. J. P. Murray, "Electromagnetic capability," Chap. 29 in *Radar Handbook*, 1st Ed., M. Skolnik (ed.), New York: McGraw-Hill, 1970.
39. E. Brookner and R. J. Bonneau, "Spectra of rounded trapezoidal pulses having an AM/PM modulation and its application to out-of-band radiation," *Microwave J*., vol. 16, pp. 49–51, December 1983.
40. A. A. Acker, "Eliminating transmitted clutter in doppler radar systems," *Microwave J*., vol. 18, No. 11, pp. 47–50, November 1975.
41. M. T. Ngo, V. Gregers-Hansen, and H. R. Ward, "Transmitter noise compensation—A signal processing technique for improving clutter suppression," *Proc. 2006 IEEE Conference on Radar*, 24–27 April 2006, pp. 668–672.
42. T. E. Vingst, D. R. Carter, J. A. Eshleman, and J. M. Pawlikowski, "High-power gridded tubes—1972," *Proc. IEEE*, vol. 61, pp. 357–381, March 1973.
43. R. S. Symons, "Tubes: Still vital after all these years," *IEEE Spectrum*, vol. 35, pp. 52–63, April 1998.
44. V. L. Granatstein, R. K. Parker, and C. M. Armstrong, "Vacuum electronics at the dawn of the twenty-first century," *Proc. IEEE*, vol. 87, pp. 702–718, May 1999.
45. R. S. Symons, "The constant efficiency amplifier," *NAB Broadcast Engr. Conf. Proc*., 1977, pp. 523–530.
46. R. S. Symons et al., "The constant efficiency amplifier—A progress report," presented at NAB Broadcast Engr. Conf. Proc., 1998.
47. L. Sivan, "The modulator," Chap. 9 in *Microwave Tube Transmitters*, London: Chapman & Hall, 1994.
48. T. A. Weil, "Transmitters," Chap. 4 in *Radar Handbook*, 2nd Ed., New York: McGraw-Hill, 1990, Sec. 4.8, "Pulse modulators."
49. T. A. Weil, "Transmitters," Chap. 4 in *Radar Handbook*, 2nd Ed., New York: McGraw-Hill 1990, pp. 4.13 to 4.14.
50. M. P. J. Gaudreau et al., "Solid state radar modulators," presented at 24th International Power Modulator Symposium, June 2000. (Available from Diversified Technologies, Inc., www.divtecs.com.)
51. M. Gaudreau et al., "Solid-state upgrade for the COBRA JUDY S-band phased array radar," presented at 2006 IEEE Radar Conference. (Available from DTI Internet site www.divtecs.com.)
52. M. Gaudreau et al., "High performance, solid-state high voltage radar modulators," presented at 2005 Pulsed Power Conference. (Available from DTI Internet site www.divtecs.com.)
53. K. J. Lee, C. Corson, and G. Mols, "A 250 kW solid-state AN/SPS-40 radar transmitter," *Microwave J*, vol. 26, pp. 93–105, July 1983.
54. J. W. Taylor, Jr. and G. Brunnis, "Design of a new airport surveillance radar (ASR-9)," *Proc. IEEE*, vol. 73, pp. 284–289, February 1985.
55. E. L. Cole et al., "ASR-12: A next generation solid state air traffic control radar," *Proc. for RADARCON 98, 1998 IEEE Radar Conference*, 12–14 May 1998, pp. 9–14.

Chapter 11
Solid-State Transmitters

Michael T. Borkowski

Raytheon Company

11.1 INTRODUCTION

For commercial applications, the transistor has all but replaced vacuum tube technology in transmitters operating at VHF and below. Since the 1980s, the power output capability among various solid-state technologies has increased to the point where they are actively pursued as replacements for some vacuum electronics in radar transmitters; this is not, however, a universally attractive solution. The transition from high-power klystrons, traveling wave tubes (TWTs), crossed-field amplifiers (CFAs), and magnetrons, to solid-state electronics has actually been very gradual because the power output of individual solid-state devices is quite limited compared to typical radar requirements. Nevertheless, transmitter designers have learned that the required higher power levels for radar transmitters can be achieved with a solid-state technology because transistors and transistor amplifier modules can be readily combined in parallel to achieve a composite higher equivalent power output. As depicted in Figure 11.1, this design attribute helps to extend the solid-state performance envelope well into the region that had previously been dominated by only vacuum electronics.[1,2] It is not the intent of this chapter to delineate the relative merits of these sometimes competing technologies, but rather to describe the limits, design practices, and characteristics of the solid-state technology for use in the common radar frequency ranges. The advantages of solid-state technologies will be described; some of the key semiconductor technologies and devices will be discussed; and some examples of solid-state component and transmitter design will be presented.

11.2 ADVANTAGES OF SOLID STATE

Although the gap in performance capabilities between the solid-state and vacuum electronics technologies can be wide, there still exist relevant trades involving cost, maintainability, and reliability, and this design tradespace can be very complicated. Some point to the continued maturation of vacuum electronics[3] and suggest that both vacuum tubes and solid-state devices will be appealing in high performance radars for many years to come. Others note still that the best value in electronic equipment is provided when the "appropriate technology"[4] is applied to affordable military electronics, recognizing that tubes and solid state may remain as complementary design solutions

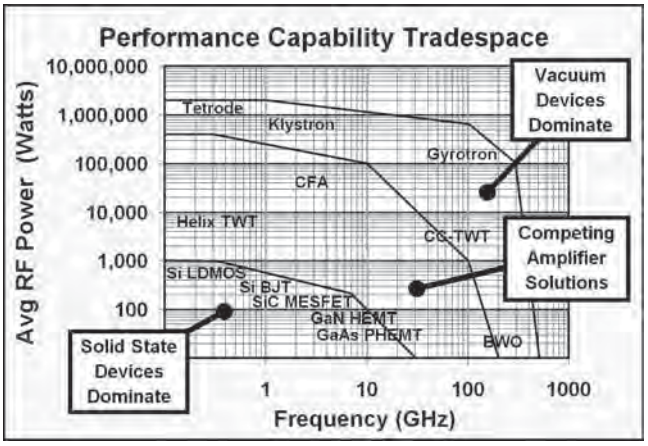


FIGURE 11.1 By combining the outputs of thousands of transistor amplifiers, the cumulative average power output of solid-state technologies can effectively compete with the performance capabilities of vacuum tube technology, as shown in the center overlap region where competing amplifier solutions coexist.

for future system requirements. For example, microwave power tubes continue to provide significantly higher power output and efficiencies than solid-state power amplifiers for high performance millimeter-wave radars.[5] Compared with tubes, solid-state devices offer the following advantages:

- No hot cathodes are required; therefore, there is no warmup delay, no wasted heater power, and virtually no limit on transistor operating life. Under certain operating conditions, the prediction of the median time to failure (MTTF) for some RF transistors can exceed 1,000 years.
- Transistor amplifiers operate at much lower voltages; therefore, power supply voltages are on the order of volts rather than kilovolts to avoid the need for large spacings, oil filling, or encapsulation. Compared with a high-voltage power supply, a low-voltage supply uses fewer nonstandard parts and is generally less expensive.
- Transmitters designed with solid-state devices exhibit improved meantime between failures (MTBF) in comparison with tube-type transmitters. Amplifier module MTBFs greater than 500,000 hours have been extrapolated from accelerated life testing. A factor of 4 improvement in transmitter system MTBF has been reported for an S band solid-state transmitter replacing a klystron transmitter.[6]
- Graceful degradation of system performance occurs when individual modules fail. Power output degrades by 20*log (1–β) as devices fail, where β is the fraction of failed devices.[7] This results because a large number of solid-state devices must be combined to provide the power for a radar transmitter, and they are easily combined in ways that degrade gracefully when individual units fail.
- The ability to demonstrate wide bandwidth is a significant characteristic of solid-state devices. While high-power microwave radar tubes can achieve 10 to 20% bandwidth, solid-state transmitter modules can achieve up to 50% bandwidth or more with acceptable efficiency.

- Flexibility can be realized. A module with both transmit and receive path amplifiers (T/R module) can be associated with every antenna element in phased array systems. RF distribution losses that normally occur in a tube-powered system between a point-source tube amplifier and the face of the array are thus eliminated. In addition, phase shifting for beam steering can be implemented at low power levels on the input feed side of an active array module; this avoids the high-power losses of the phase shifters at the radiating elements and raises overall efficiency. Also, peak RF power levels at any point are relatively low because the outputs are combined only in space. Furthermore, amplitude tapering can be accomplished by turning off or attenuating individual active array amplifiers. For phased array systems with modest power levels, the solid-state solution offers advantages that make it attractive as the basis for a radar transmitter.

The general replacement of high-power microwave tubes by solid-state devices has not been straightforward. Attempts to replace existing tube-type transmitters with a solid-state retrofit have been hindered by the requirement to be a form, fit, functional replacement for the incumbent hardware. Radar transmit waveforms that previously had been architected to make optimum use of the high peak power and low duty cycle capability of the tube no longer favor the solid-state transmitter. A low duty cycle environment is not the most cost-effective solution for solid-state devices because transistors exhibit much shorter thermal time constants than the replacement tube and are more efficiently operated using a lower composite peak power at a higher duty cycle. As an example of the dilemma, an L-band microwave transistor that is capable of perhaps 50 watts (W) average power cannot provide much more than 300 watts of peak power without overheating during the pulse. The short pulse lengths and low duty cycles typical of older tube-type radars thus make very inefficient use of the average power capabilities of microwave transistors. To replace the old, well-proven 5J26 L-band magnetron that develops 500 W of average RF power at 0.1% (typical) duty cycle would require 2500 to 5000 of the 50-watt transistors just described. However, with a 10% duty cycle the 500-watt average power requirement could be provided by only 25 to 50 of the 50-watt transistors. In other words, microwave transistors are much more cost-effective when the required radar system average power can be provided by a lower peak power at a higher duty cycle. As a result, there have been relatively few direct replacements of older low duty cycle transmitters by solid-state transmitters. Some initiatives, such as the solid-state AN/SPS-40 replacement that was motivated by the attractive reliability, maintainability, and availability characteristics of a modular solid-state system, have not seen the success once envisioned due to the acquisition cost of solid-state replacement transmitters. For new radar systems, system designers have been motivated by these considerations to choose as high a duty cycle as possible, both to reduce the peak power required and to permit using solid-state devices at a reasonable cost.

The decision to use a high transmitter duty cycle, however, has significant impact on the rest of the radar system. Operation at a high duty cycle generally requires the use of pulse compression to provide the desired unambiguous range coverage together with reasonably small range resolution. Other consequences follow in turn: the wide transmitted pulse used with pulse compression blinds the radar at short ranges, so a "fill-in" pulse must also be transmitted and processed. To prevent points of strong clutter from masking small moving targets, the signal processor must achieve low pulse compression time sidelobes and high clutter cancellation ratio. As a result, it is much easier to design a solid-state transmitter as part of a new system than it is to retrofit one into an old system that usually does not have all these features.

The use of solid-state does not eliminate all the problems of transmitter design. The RF combining networks must be designed with great care and skill to minimize combining losses in order to keep transmitter efficiency high. Suitable isolation from excessive voltage-standing-wave ratio (VSWR) must be provided to protect the microwave transistors from undesired operational stresses, and their harmonic power output must be properly filtered to meet MIL-STD-469 and other specifications on RF spectrum quality. Also, just as in tube-type transmitters, energy management is still crucial. Each dc power supply must have a capacitor bank large enough to supply the energy drawn by its solid-state modules during an entire pulse, and each power supply must recharge its capacitor bank smoothly between pulses without drawing an excessive current surge from the power line.

As a result of unavoidable losses in combining the outputs of many solid-state devices, it is especially tempting to avoid combining before radiating, since combining in space is essentially lossless. For this reason, many solid-state transmitters consist of amplifier modules that feed either rows, columns, or single elements of an array antenna. Especially in the last-named case, it is necessary to build the modules (and usually their power supplies) into the array structure. Generally, solid-state devices or modules are combined in one of three fundamental configurations to generate the required transmitter power levels. Figure 11.2 shows that this may involve either the combination of amplifier outputs to a single port to feed a mechanically rotating antenna or some combination of electronic phase steering and amplification distributed among many fixed elements of a planar two-dimensional array.

Because of the large number of individual modules in a typical solid-state transmitter, failure of an individual or a few modules has little effect on overall transmitter performance. The module outputs add as voltage vectors, so the loss of 20% of the modules, for example, results in a reduction to 80% of voltage output, which is 64% of power output. Even this is only a 2-dB reduction (the difference between 64 and 80% of the power ends up in the combiner loads or in sidelobes if the combining is in space). As a result of this "graceful degradation," overall reliability of solid-state transmitters is very high even if maintenance is delayed until convenient scheduled periods; however, this advantage should not be abused. Consider a case where 20% of 1000 modules are allowed to fail before output power falls below requirements, and assume that maintenance occurs at scheduled three-month intervals. In this case, module MTBF need only be 22,000 h to provide 90% confidence that the transmitter

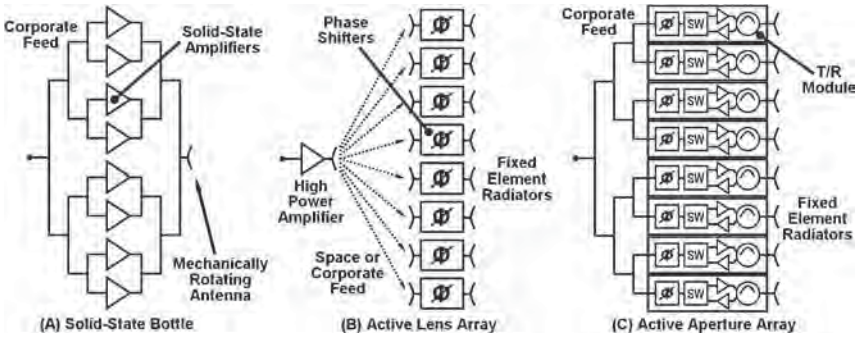


FIGURE 11.2 Common solid-state transmitter configurations may combine many amplifiers in parallel to a single antenna port (*a*), or may use phase-shift elements to electronically steer a beam (*b*), or may utilize transmit/receive modules with phase-shift capability at every element to steer a beam (*c*).

will not "fail" in less than three months; however, the cost of replacement modules and labor would be very unattractive because nearly 40% of the transmitter would have to be replaced every year. Higher MTBFs are thus essential to ensure that the transmitter is not only available but also affordable. Fortunately, solid-state module reliability has proven to be even better than the MIL-HDBK-217 predictions. AN/FPS-115 (PAVE PAWS), for example, has grown to 141,000 hours, which is 2.3 times the predicted value. This includes the actual T/R module MTBF, along with the receiver transmit/receiver (T/R) switches and phase shifters as well as the power amplifiers. In fact, MTBF for the output power transistors measures better than 1.1 million hours.

11.3 SOLID-STATE DEVICES

Although the RF power-generating capability of single transistors is small with respect to the overall peak and average power requirements of a radar transmitter, transistors are used quite effectively by combining the outputs of many identical solid-state amplifiers. The power output level from a particular device is a function of not only the chosen technology, but also the frequency and other conditions, such as pulse width, duty cycle, ambient temperature, operating voltage, and the presented load impedance.

Technologies and Construction. Semiconducting materials used in the fabrication of transistors are considered to be those materials that are typically neither conductors nor insulators. The charge carrying properties of these semiconducting materials can be modified dramatically through the substitution of minute amounts of impurity ions or through crystal lattice defects, either of which act to modulate the flow of electrons. Semiconductor materials from which transistors are fabricated for use in solid-state radar transmitters have generally been either silicon or one of the so-called compound semiconductors, such as gallium arsenide (GaAs), indium phosphide (InP), silicon carbide (SiC), gallium nitride (GaN), or silicon germanium (SiGe). Semiconductors like silicon or gallium arsenide have found early wide acceptance because it has proven practical to control their crystal lattice defects accurately and repeatably during transistor manufacturing. Some semiconductors, such as gallium nitride (GaN) or silicon carbide (SiC), are referred to as wide bandgap semiconductors. Semiconductors that exhibit large bandgap values are especially capable of producing very high output power levels with acceptable gain at the frequencies used in most radar applications.

Transistors are three-terminal devices and are classified as either bipolar or unipolar. Figure 11.3 helps to portray the construction differences among common microwave three-terminal devices, and this figure is referenced multiple times in succeeding sections. The bipolar junction transistor (BJT) is so named because the conduction path through the transistor makes use of both majority and minority charge carriers to establish current flow in the semiconductor. It is a current-controlled device with the collector current modulated by the current flowing between the base-emitter junction. This compares to the operation of a field effect transistor (FET), or a unipolar device, where charge is carried with only one type of charge carrier. The remaining transistor constructions in Figure 11.3 are all variants of a FET. An external voltage, applied to the gate terminal of a FET, controls the width of the depletion region below the gate terminal. As the width of the depletion region is varied, so too is the equivalent resistance between

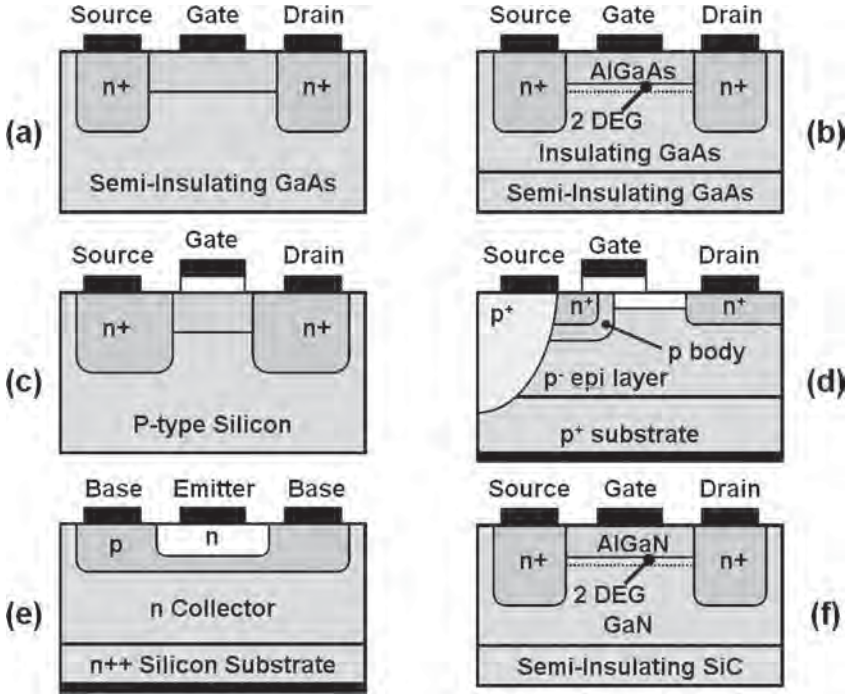


FIGURE 11.3 Transistors are three-terminal semiconductor devices that allow a small voltage or current to control a larger voltage or current. Some cross sections of common transistor types used in the design of radar transmitters are the (*a*) GaAs MESFET, (*b*) GaAs PHEMT, (*c*) Silicon MOSFET, (*d*) Silicon LDMOS FET, (*e*) Silicon BJT, and (*f*) GaN HEMT on SiC Substrate.

the drain and source contacts, allowing the current flowing between the drain and source to be modulated accordingly; hence, FETs are referred to as *voltage controlled devices*. There exist numerous FET variants due to sometimes subtle construction or material differences. Among these are the MOSFET (Metal-Oxide-Semiconductor FET), MESFET (Metal Semiconductor FET), and HFET (Heterostructure FET). The common HFET devices are referred to as the HEMT (High Electron Mobility Transistor) and PHEMT (Pseudomorphic High Electron Mobility Transistor).

To be useful in a radar transmitter amplifier, the transistor must be capable of operating at the appropriate high frequency with good efficiency, while demonstrating useful power gain with adequate thermal management properties to ensure high reliability. There is not one transistor type or one semiconductor material that is universally useful across all the common radar bands from UHF through W band. In fact, among the radar bands, there is often a different dominant device type, along with its attendant design and fabrication methodologies, that offers the optimum performance for that band.

A first order approximation[8] of the power output from a single-stage solid-state amplifier using a FET as the semiconductor device is given by the relationship

$$P_{\mathrm{RFMAX}} = I_{\mathrm{MAX}} * (V_{\mathrm{DGB}} - |V_P| - V_K)/8 \tag{11.1}$$

In this relationship, I_{MAX} is the maximum open channel current, V_{DGB} is the gate-drain breakdown voltage, V_P is the pinchoff voltage, and V_K is the knee voltage. These transistor parameters define the transfer characteristics in the I-V (current-voltage) plane, and the boundaries of the I-V plane define the peak power performance envelope of the transistor. Also, there is an optimum load impedance that will maximize the power output that can be delivered from an amplifier; and to a first order estimate, that load impedance is represented by the line that transversely cuts across the I-V plane from the region of the breakdown voltage to the region of the knee voltage, as shown in Figure 11.4. The ability of a transistor to demonstrate gain at high frequencies is impacted by the mobility and saturated velocity of charge carriers in the semiconductor. The ability of a transistor to demonstrate high power output is impacted by the breakdown voltage, the current capability, and the knee voltage of the transistor.

Silicon device types cost-effectively satisfy the requirements of reliability, electrical performance, packaging, cooling, availability, and maintainability at lower radar-band frequencies, typically UHF, L band, and into S band. These devices are usually manufactured as discretely packaged transistors and require external impedance-matching circuitry in order to function appropriately in an amplifier. High performance transistors at higher frequencies than S band are usually built using compound semicondctors. Such transistors can result in high cutoff frequencies and demonstrate gain at frequencies much higher than silicon. For example, electrons in gallium arsenide (GaAs) travel approximately twice as fast as they do in silicon. It has a higher saturated electron velocity and higher electron mobility, allowing it to function at frequencies into the W band. GaAs transistors generate less noise than silicon devices when operated at high frequency so they also make superior low-noise amplifiers. A key attribute that makes GaAs an attractive technology is that the GaAs FET can be fully integrated with the passive circuitry that is necessary to provide the biasing, loading, filtering, and switching functions that are necessary for multistage T/R module designs. Unlike the silicon power transistors, the GaAs FET and its associated

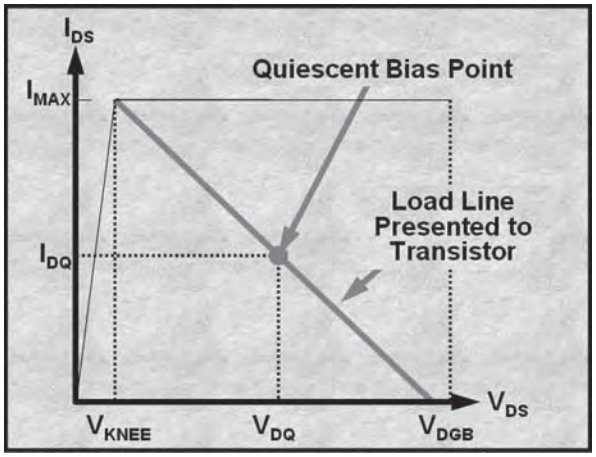


FIGURE 11.4 Typical transistor current-voltage continuum (I-V plane) showing key FET dc performance limits with optimum load line for power output shown. Higher power output is achieved when the maximum channel current (I_{MAX}) and the breakdown voltage (V_{DGB}) are both increased. Optimum amplifier design places the load line as indicated.

TABLE 11.1 Key Semiconductor Characteristics of the Primary Semiconductors Used for Power Generation in Solid-state Transmitters (The high saturated velocity, breakdown field, and thermal conductivity of SiC and GaN make them attractive for high-power-amplifier applications.)

	Units	Silicon	GaAs	InP	SiC	GaN
Bandgap Energy	eV	1.1	1.4	1.3	3.2	3.4
RF Power Density	W/mm	0.6–0.8	0.8–1.8	0.2–0.4	2.0–4.0	3.0–10.0
Dielectric Constant	—	11.8	12.8	12.5	9.7	9.0
Breakdown Field	10^6 V/cm	0.6	0.7	0.5	2.5	3.5
Thermal Conductivity	W/m°C	130	46	68	370	170
Electron Mobility	cm^2/Vsec	700	4700	5400	600	1600
Saturated Velocity	10^7 cm/sec	1.0	2.0	0.9	2.0	2.5

batch-processed monolithic microwave integrated circuitry (MMIC) fabrication technology allow for circuit functions to be processed into very, very small, conveniently packaged chips. The wide bandgap semiconductors, such as SiC and GaN, are also compatible with MMIC processing but are also capable of very high power output levels. These semiconductors have material properties that lead to high breakdown voltage with commensurately high channel currents—an order of magnitude higher power output than GaAs (Table 11.1).

For the upper end of the solid-state microwave spectrum, i.e., the millimeter-wave range, the single-port microwave diode can be used as a low-power oscillator. Unfortunately, the power output and efficiency of these devices are, in general, very low; in fact, the efficiency is significantly lower than that of their tube counterparts. However, CW and pulsed power output are attainable up to 300 GHz.

Peak and Average Power Limitations. A first-order limit on the RF power output capability of a transistor is its breakdown voltage and maximum current handling capability. Within that limit, the maximum practical level of power output that can be obtained from a single transistor over a given bandwidth is governed by the thermal dissipation limit of the device. As devices become larger and the dissipative heat flux from the top surface of the transistor chip to the bottom layer of the transistor chip increases, the junction temperature increases to the point where the transistor becomes thermally limited. Regardless of the semiconductor used, the electrical performance and operating lifetime degrade at increasingly higher temperatures.

There is a composite thermal time constant associated with the numerous thermally resistive layers between the transistor junction and the heat sink or cold plate to which the device is attached. This occurs because each layer (semiconductor, ceramic substrate, metal base, etc.) exhibits both a thermal resistance and a thermal capacitance. Then, there exists an equivalent thermal time constant (τ) for each packaging material layer. This thermal time constant has been approximateded[9] as

$$\tau = 0.4053 * (F^2 \rho C / K_{TH}) \tag{11.2}$$

where F is thickness (cm), ρ is density (gm/cc), C is specific heat (Wsec/gm°C), and K_{TH} is thermal conductivity (W/cm°C). For example, Figure 11.5 shows that when the pulse width and duty cycle for a given GaAs transistor is increased from 20 μs and 10% to 3000 μs and 25%, respectively, there is a 70°C increase in the overall junction temperature. Although the transistor may operate reliably at a desired output power level for the shorter pulse width, it would suffer from a decrease in long-term reliability if operated at the same power level for the longer pulse width. Thus, if required to

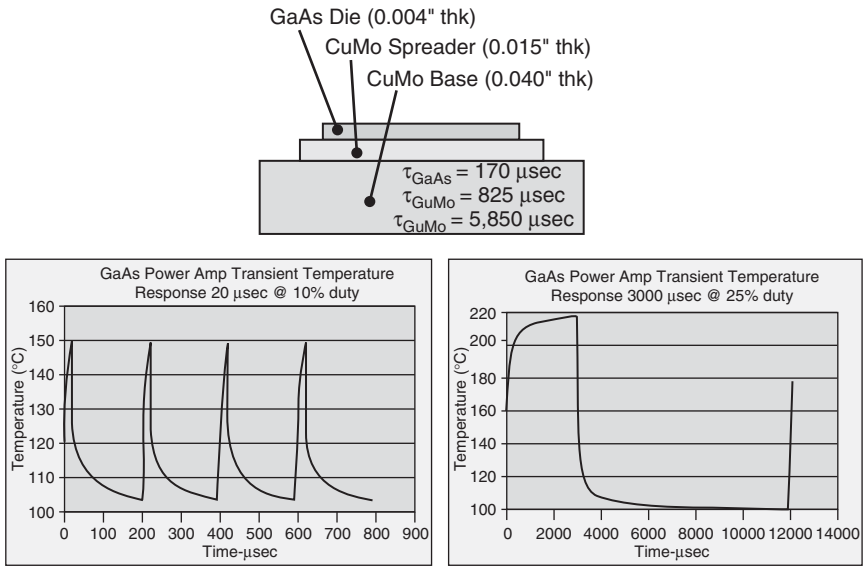


FIGURE 11.5 One limit of transistor capability is determined by the maximum junction temperature, which in turn is determined by the thermal time constant, and this results in very different capabilities as a function of operating pulse width and duty cycle.

maintain long-term reliability at the longer pulse width, the dissipation in the transistor would have to be reduced to bring the junction temperature down to an acceptable level. De-rating the inherent short-pulse capability by lowering the dissipated power in the transistor, perhaps by reducing the operating voltage and power output, is one method of achieving the desired reliability. Another method may involve the reduction in ambient temperature with the use of chilled fluid in the amplifier heat-sink. These are not always practical solutions and one finds that the layout of the transistor itself is often optimized for a particular pulse width and duty cycle in order to achieve the optimum performance and reliability at the lowest operating temperature.

The active transistor area on the surface of the chip (the chip is sometimes called the *die*) is typically divided into manageable units (cells), where the cell size is often optimized for a particular application or range of applications. In addition to frequency considerations, pulse width and duty cycle or, as a result, the peak and average dissipated power, are the parameters that determine the cell size and arrangement of cells on a chip. The ultimate operating junction temperature of the transistor is largely dependent on the transient heating that will be encountered and the layout and area of the individual cells. For devices that are designed to operate for long pulses or CW, an increase in the average power capability of the transistor can be achieved by dividing the active area of a transistor into small, thermally isolated cell areas.

Since the overall thermal time constant for a typical power transistor die itself may be on the order of 100–200 μs, the tradeoff between peak and average power versus device size can be significant for solid-state radars using pulse compression with pulse widths in the 10 to 1000 μs range. As an example, the thermal time constant of a silicon die with a thickness of 5 mils is approximately 90 μs whereas a gallium arsenide die with a thickness of 4 mils is approximately 170 μs. Thus, for an operating pulse width

representative of a solid-state radar with pulse compression (~300 μs), the temperature rise across the silicon die has reached 96% of its steady-state value, but for an operating pulse width representative of a shorter range fire control radar (~20 μs), the temperature rise across the silicon die has only reached 20% of its steady-state value. If the voltage and current threshold of the transistor has not been reached, the shorter pulse width operation could allow for significantly larger power capability to be demonstrated. Usually, a very detailed thermal analysis using the finite element method is required to quantify these relationships during transistor and amplifier design.

Background and descriptions of the common three-terminal device types and their associated technologies as utilized for the common radar bands are described in the following sections.

Silicon Bipolar Junction Transistor. The silicon bipolar junction transistor (BJT) was the earliest of the microwave power devices and found its way into tube replacement transmitters and phased array applications starting in the late 1970s. At lower frequencies, especially below 3 GHz, the Si BJT has been shown to be capable of very high power levels for transistors. Amplifier design is realizable for frequencies up through S band, where the tradeoff between device performance and overall system cost begins to reach a point of diminishing returns. The silicon bipolar transistor technology is now very mature, but the demand for these high performance devices is low because the production quantities required for radar systems is small relative to commercial silicon electronic products. Thus, there tends to be a small number of manufacturers who provide quality devices for use in amplifier designs.

Silicon-based microwave power transistors can actually be considered hybrid microelectronic circuits and are generally single-chip or multichip transistors combined in parallel within a flanged hermetic package. Some form of internal impedance prematching circuitry is often included in order to preserve the intrinsic bandwidth of the semiconductor chip and to make the task of external impedance matching easier. The internal matching also increases the terminal impedances of the packaged device to a level where the component losses of the circuitry external to the transistor become less critical. Figure 11.6 is an example of a 230-watt internally matched power transistor hybrid using the Si BJT semiconductor technology. It shows the transistor die, along with the capacitors and wires that are used as low-pass and high-pass impedance-matching components to achieve an acceptable level of impedance prematching.

The microwave power Si BJT is invariably an NPN structure (Figure 11.3*e*) with a vertical diffusion profile; i.e., the collector contact forms the bottom layer of the chip. The P-type base region has been diffused or implanted into the collector, the N-type emitter has been diffused or implanted into the base, and both base and emitter regions are accessible from the top surface of the chip. The collector region consists of an N-doped, low-resistivity epitaxial layer that is grown on a very low resistivity silicon substrate. The characteristics of the epitaxial layer, i.e., thickness and resistivity, can determine the upper limit of performance of the device in terms of ruggedness, efficiency, and saturated power output.

The fundamental limitation on high-frequency Si BJT performance is the overall collector-to-emitter delay time. If a signal is introduced to either the base or the emitter, four separate regions of attenuation or time delay are encountered: the emitter-base junction capacity charging time, the base transit time, the collector depletion-layer transmit time, and the collector capacitance-resistance charging time. High-frequency transistor design is concerned with optimizing the physical parameters that contribute to the time-delay components.[10]

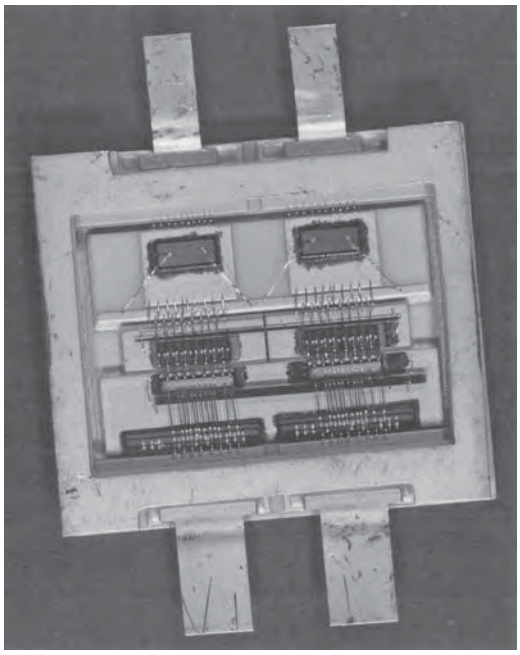


FIGURE 11.6 A 230-watt L-band long-pulse and high-duty-cycle silicon bipolar power transistor in a custom hermetic, dual-leaded, low-inductance package has an overall footprint of 0.40" × 0.45." (*Photograph courtesy of Raytheon Company*)

The design challenge for high-power Si BJTs is to maintain a uniform high current density over a large emitter area with a minimum temperature rise. High-frequency devices require shallow, narrow, high-resistance base regions under the emitter region, causing most of the current carried in the device to be crowded along the periphery of the emitter. Thus, in order to maximize the current-handling capability of the device and, hence, the power output capability of the device, the emitter periphery is maximized. Because the capacitance of the collector-base junction appears as a deleterious parasitic electrical component, the emitter-periphery to base-area ratio, or Ep/Ba, is maximized where possible. Generally, higher-frequency devices exhibit higher Ep/Ba ratios; and to obtain a high Ep/Ba ratio very fine line geometries are required, where the term *geometry* refers to the surface construction details of the transistor dice.

Silicon LDMOS FET. The silicon Laterally Diffused Metal-Oxide Semiconductor (LDMOS) transistor is beginning to supersede the silicon power BJT as a replacement device, especially at the VHF, UHF, and L-band frequencies. In particular, the commercial communications industry has found that the Si LDMOS FET dominates as a cell phone base-station power amplifier because of the higher gain, linearity, and efficiency that it demonstrates compared to the silicon BJT. Although it is a FET, its construction characteristics, packaging, and design challenges are very similar to the design challenges of the Si BJT.

The silicon LDMOS FET (Figure 11.3*d*) is processed on p+ material with a lightly doped p-type epitaxial layer, and just like the silicon BJT, multiple impurity implants form the various junctions. It is still considered a slower device than other semiconductors, such as GaAs, because the mobility in silicon MOSFET channels is relatively low. Although the bulk mobility of silicon is lower than GaAs, it does not preclude the silicon LDMOS FET as a high frequency power transistor. The continuous process fabrication improvement in the silicon CMOS industry has resulted in sub-micron production transistor feature sizes, and the smaller feature sizes allow for a compensated increase in higher frequency operation; that is, it can exhibit usable gain into S band. In particular, the LDMOS structure enables a short channel as a result of the lateral diffusion of the p-type implant. The resulting short channel contributes to improved high frequency response in spite of the lower mobility of silicon. The measured breakdown voltages can be in excess of 100 V, so operation at higher voltages is possible, or conversely, a higher level of margin in ruggedness can be achieved for a given operating voltage; the latter is a key advantage for high-reliability power amplifier applications.

The underside of the LDMOS transistor dice is the source connection; thus, the chip can be mounted directly to a metal package base. This is unlike the Si BJT where the high voltage collector contact is the underside of the chip. As a result of not having to electrically isolate the underside of the LDMOS chip, there is no need to use the potentially toxic beryllium oxide based packages that permeate the Si BJT product lines. The lower source inductance achieved with direct attachment to the metal flange of a package base enables higher gains to be demonstrated for comparable Si BJT power levels at frequencies below 2 GHz, but these devices are presently not attractive at frequencies above S band.

A primary advantage of the LDMOS device is thermal stability. The drain current has a negative temperature coefficient; therefore, the LDMOS FET is not susceptible to thermal runaway and does not require the amount of gain-degrading resistive emitter ballasting that is commonly required in a Si BJT to help normalize junction temperatures. A more thermally stable device allows for the more efficient power combining of transistor cells within a package. This contributes to a lower performance sensitivity to load mismatch—a problem that has complicated the design process for the Si BJT. Figure 11.7 and Figure 11.8 summarize the performance envelope for commercially available silicon bipolar junction transistors and silicon LDMOS FETs for given transmit waveforms.[11–14]

GaAs PHEMT. The GaAs Pseudomorphic High Electron Mobility Transistor (PHEMT) is actually a heterostructure material (Figure 11.3*b*) consisting of slightly strain-mismatched crystal layers. A very simplified description portrays an AlGaAs layer over an InGaAs channel on a GaAs substrate forming a high quality two-dimensional electron gas layer, often referred to as the 2DEG. This 2DEG exhibits superior electron transport properties resulting in a very confined channel with fewer opportunities for charge carrier collisions. This allows for a very high-quality transistor that can be made with useful gain beyond W band. Higher mobility and electron velocity can be engineered by increasing the percentage content of indium in the channel of the FET. This can be accomplished up to a point, where beyond approximately 25% indium content, the lattice strain differential results in degrading performance and reliability. These techniques can result in transistors with bigger bandgap differences than otherwise possible for the chosen materials. The fabrication of these transistors employs the use of advanced semiconductor processing such as Molecular Beam Epitaxy (MBE) or Molecular Organic Chemical Vapor Deposition

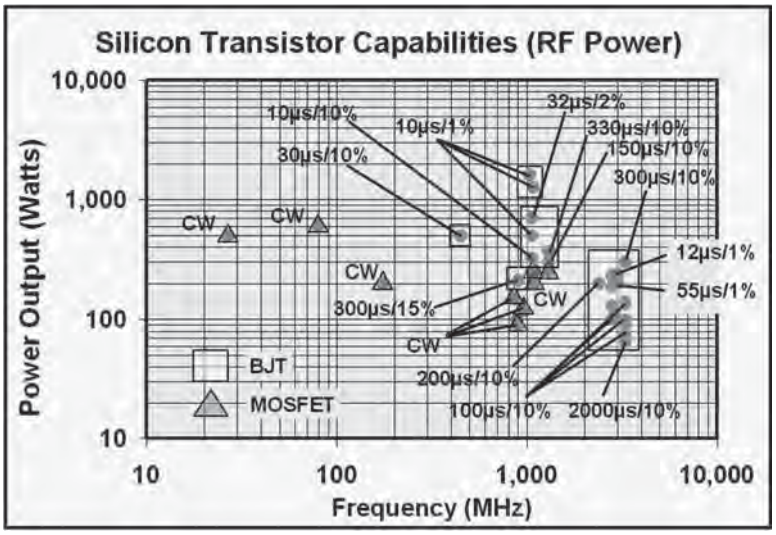


FIGURE 11.7 Performance space of commercially available silicon power transistors

(MOCVD) fabrication techniques to achieve the high performance characteristics. These are capital-intensive semiconductor processing steps that are required in order to achieve the high quality channel characteristics that define a microwave or millimeter-wave power transistor. Optimizing the molecular content of the FET channel for best performance is sometimes referred to as *bandgap engineering*. The challenge for the device engineer is to develop a transistor that supports the highest voltage/current operation while demonstrating the best high frequency gain.

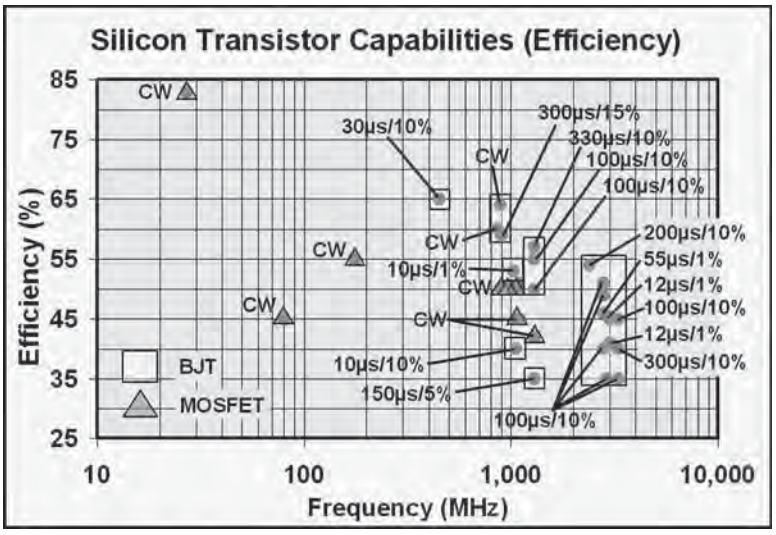


FIGURE 11.8 Performance space of commercially available silicon power transistors

The complication in power transistor design extends beyond the exotic material fabrication technologies that are used to define the basic FET. Tailored construction techniques are used to control the electric field intensity and improve the breakdown voltage; enhancements such as the field plate,[15,16] double gate recess,[17] or automatic etch stop layers[18] are fabrication and design techniques that are used to optimize the PHEMT performance for a given operating frequency range to bring higher value, performance, or reliability to the semiconductor fabrication process.

The fundamental three-fingered FET structure (drain-gate-source), shown as a GaAs PHEMT FET in cross-section in Figure 11.9, has power performance limitations that are mitigated by employing a structured design approach for achieving high power-output levels. As shown in Figure 11.10, the smallest physical construction dimension of the gate electrode is called the *gate length*; the longer of the dimensions is called the *gate width*. The current carrying capability (and hence the power capability) of the FET is increased as the gate width is increased. There is a limit as to how long the gate width can be increased before phase differential and signal attenuation along the longer dimension of the gate width begin to have a deleterious impact on amplifier performance. In practice, the maximum gate width can be found to approach approximately 400 μm, 150 μm, or 60 μm for S band, X band, or K_a band frequencies, respectively.

With a limitation on maximum gate width, additional current and hence power capability can only be achieved by combining multiple gate electrodes in parallel. Gate electrodes, or *fingers* as they are often referred to, are generally grouped in logically convenient substructures (cells) that are stepped and repeated to form a symmetric corporate hierarchy of combined transistors. The outputs of all fingers are required to be combined in phase and then impedance-matched to the appropriate level. An industry-wide figure-of-merit for capability of the semiconductor and the unit FET is the power output density, and this is given in units of watts/mm of total FET gate width or gate periphery. For operating voltages of 7–10 volts, a normalized power output density of 0.6–0.8 W/mm should be expected; for more advanced GaAs PHEMT structures operating at 11–28 volts, one can expect to deliver 1.1–2.0 W/mm of normalized power output density. Thus, to achieve a power level of 20 watts at 10 GHz, when operating from 15 volts, approximately 80 gate fingers must somehow be combined in parallel to

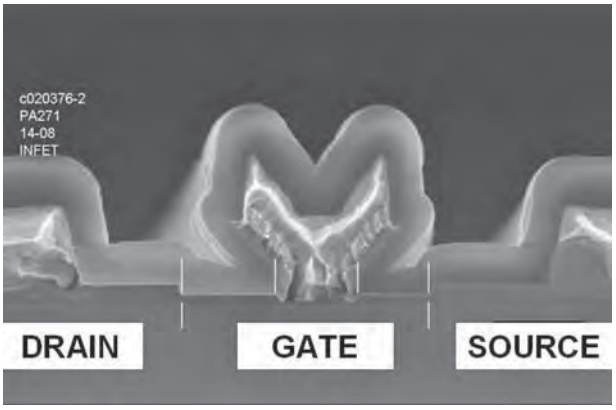


FIGURE 11.9 Cross section of a 0.25 μm double gate–recessed GaAs PHEMT transistor, showing gate, drain, and source metals (*Photograph courtesy of Raytheon Company*)

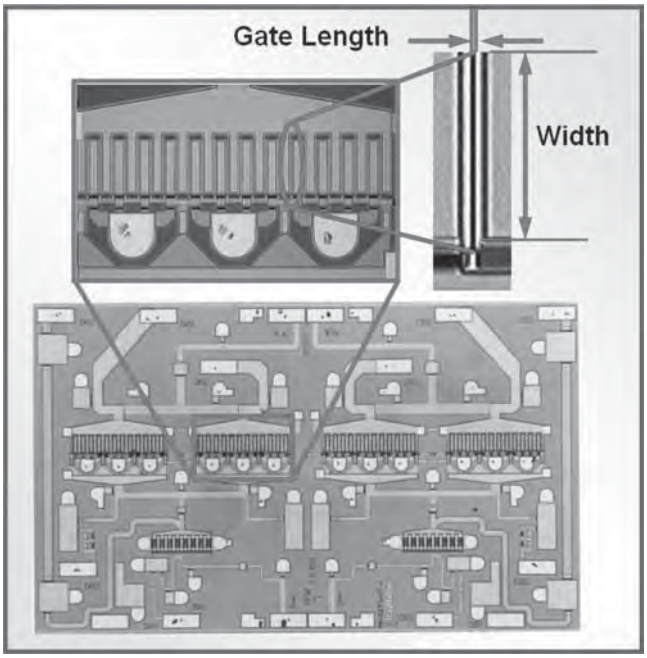


FIGURE 11.10 Typical two-stage GaAs MMIC power amplifier with insert showing multiple paralleled gate fingers in final stage unit cell (*Photograph courtesy of Raytheon Company*)

deliver the power. Higher numbers of paralleled fingers necessarily lead to decreasing input and output impedances, further complicating the ability to provide the desired impedance match over the desired bandwidth. Higher impedance transformation ratios and wider bandwidths invariably contribute to additional loss in the matching networks, whether these networks are hybrid construction or MMIC construction. Additional losses degrade the inherent power, gain, and efficiency characteristics of the intrinsic FET. The most appropriate amplifier design for a given application requires optimization of the transistor, and the variables that affect that optimization, such as unit gate length, gate width, number of gate fingers, cell construction, impedance matching circuits, and bias networks, all require detailed attention during the design of the amplifier. Excellent compilations of industry performance have been published for further insight.[19,20] These references outline state-of-the-art power output density and efficiencies of compound semiconductors from 1 through 100 GHz.

Wide Bandgap Semiconductors. Since the inception of solid-state, Zolper[21] draws historical reference to the first, second, and third generation of semiconductor materials as (1) silicon, (2) gallium arsenide or indium phosphide, and the (3) so-called wide bandgap semiconductors (WBGS). The latter are dominated by the silicon carbide (SiC) MESFET and the gallium nitride (AlGaN/GaN) heterojunction field effect transistor (HFET). The advent of the third semiconductor generation opens enormous new possibilities in the area of high power amplifiers for use in solid-state. The WBGS materials are able to produce very high power-output levels (5–20 w/mm)

from high bus voltages (25–75 volts), while maintaining transistor-like properties at higher operating temperatures than Si or GaAs.[22] They are finding application in the S band, C band, or X band frequency ranges. The interest in development is being fueled by both the military for high performance sensors and also by commercial interests for high-power wireless base-station amplifiers. In particular, the GaN HFET device demonstrates physical properties that make it useful as a high gain device with very high power output capability into the W band. The SiC MESFET will likely be competitive at the lower frequency ranges of L band through C band.[23]

The thermal conductivity of the SiC substrate is superior to GaAs by nearly an order of magnitude, and the normalized power outputs are much, much higher in the wide bandgap semiconductors than currently being achieved using GaAs at any voltage. The higher thermal conductivity of SiC enables more efficient thermal management. Coupled with the high breakdown voltage and channel current capability of the SiC MESFET, measured results of 80 watts of CW power output with an 8 dB associated level of large signal gain at 3.1 GHz from a 58 V drain supply voltage have been reported[24] from a single transistor cell.

Electron mobility in the GaN HEMT at saturated drift velocities is high enough that high gain with simultaneous high power output and high efficiencies can be achieved with voltages as low as 20 to 30 volts. With a GaN epitaxial layer processed on a SiC substrate, the current state-of-the-art for transistor performance is defined on several fronts in the semiconductor industry by the following performances: (1) pulsed power-added-efficiency (PAE) of 68% at 30 V and 10 GHz on a 1.25 mm FET, and CW power of 5 W at 30 V and 10 GHz on a single 1.25 mm transistor;[25] (2) power density of 8.6 W/mm at 40 GHz;[26] (3) less than 0.2 dB power degradation after 15,000 hours RF operation at 28 V with channel temperatures of 150°C.[27] Power-added efficiency is a circuit designer's term and is defined by

$$PAE = (P_O - P_I)/P_{\mathrm{DC}} \tag{11.3}$$

where P_O is the RF power output, P_I is the RF power input, and P_{DC} is the total dc power input.

Figure 11.11 and Figure 11.12 illustrate the advantages of GaN at 10 GHz when compared with the physical geometry of an identically sized GaAs PHEMT transistor.

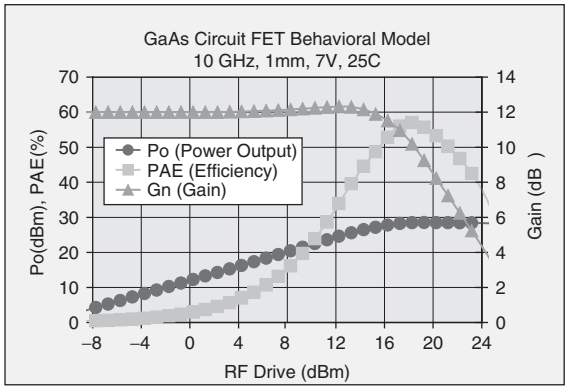


FIGURE 11.11 Typical 10-GHz performance curves for a 1 mm periphery GaAs PHEMT FET operating at +7 V using a CW (100% duty) waveform

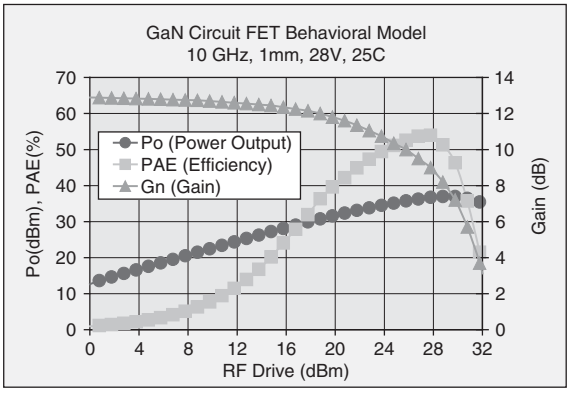


FIGURE 11.12 Typical 10-GHz performance curves for a 1 mm periphery GaN HEMT FET operating at +28 V using a CW (100% duty) waveform

In this example, a GaAs PHEMT with 1 mm of total gate periphery (Figure 11.11) is compared with an identical 1 mm GaN HEMT (Figure 11.12). These figures portray the power output (*P*o), power-added efficiency (PAE), and gain (*G*n). Performance for each is referenced to 10 GHz for CW operation. The GaAs PHEMT operates at 7 V; the GaN HEMT operates at 28 V. From these performance curves, one sees that the small signal gains are nearly identical in the 12–13 dB range, as are the large signal efficiencies, but the power output capability of the GaN HEMT is 8 dB greater than for the GaAs PHEMT of the same size.

11.4 DESIGNING FOR THE SOLID-STATE BOTTLE TRANSMITTER

Radar transmitter design invariably requires significant radiated power from the antenna in order to project to the minimum range requirement while maintaining some minimum signal-to-noise ratio on receive. The impact of the requirement for high radiated power is fundamental to the design of solid-state transmitters—high power must be achieved by combining the outputs of lower-power amplifiers in order to develop the required radiated levels. Transistor fingers are combined into MMIC amplifiers; amplifiers are combined into modules; and modules are combined into systems. Generally, the combining takes on one of two different configurations using either space-combined or corporate-combined architectures. The phased array is a common example of the space-combined configuration wherein each radiating antenna element is fed by an amplifier module and the wavefront is formed in space. The common example of the corporate-combined design is the "solid-state bottle," wherein a mechanically rotating antenna is fed from a single port, and the power at the port is the summation of the outputs of many amplifier modules. Those modules might be physically located, for example, below deck on a ship away from the antenna. Solid-state transmitter designs have been built around each of these generic forms, and the components that are required in the implementation of each share similar characteristics and devices.

In the corporate-combined solid-state bottle, high power levels are generated at a single point by combining the outputs of many power amplifier modules. In general, a power amplifier module, as shown in Figure 11.13, consists of a number of identical amplifiers that are parallel-combined and isolated from one another through the use of microwave combining and isolating techniques. Drive power for this parallel group is obtained from driver or predriver stages, using phase- and amplitude-matched mirror-imaged microwave power dividers. A circulator at the module output port is commonly used to protect the amplifier from the damaging effects of high-load VSWR, most notably from the antenna. Also, ancillary circuitry such as energy storage capacitance for pulsed operation, built-in-test (BIT) sensors, or adaptive control components may be included.

Amplifier & Module Design. Solid-state amplifiers for use in transmitter design are often referenced by their class of operation. Amplifiers are designated as operating either Class-A, -B, -AB, -C, -D, -E, -F, or -G. Class-A, -AB, -B, and -C generally refer to analog amplifiers whereas Class-D, -E, -F, and -G generally refer to switching-mode amplifiers. Each class of operation for the analog modes is defined by the manner in which the transistor is biased; each class of operation

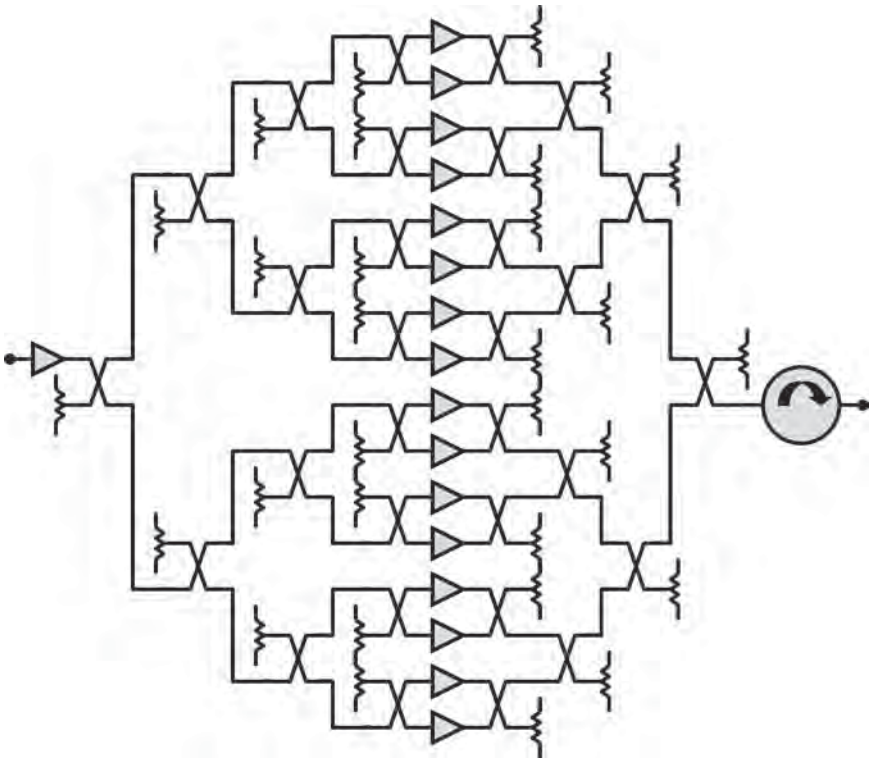


FIGURE 11.13 Solid-state power amplifier module combines many single stage amplifiers together with matched phase and amplitude using resistively isolating combining techniques

for the switching modes is defined by the manner in which the transistor is biased and how the waveshape of current and voltage is manipulated. For example, current swing in an amplifier that is biased Class-A replicates exactly the input signal up to the point where the transistor voltage and current limits are reached. In practice, Class-A amplifiers are the most linear as well as the least efficient. High dynamic-range linear receive amplifiers are biased Class-A, and audio amplifiers may be also be biased Class-A to preserve the linearity of the input signal. Class-B amplifiers are biased such that conducting current in the transistor flows for exactly one half of the input signal voltage swing. Push-pull amplifiers may be biased in this fashion such that one transistor operates over the positive input signal swing, and the second transistor operates over the negative input signal swing. Higher efficiency but higher distortion is experienced when compared with a Class-A design. The Class-AB operated amplifier is biased just above 50% conduction using a trickle quiescent current and is also commonly used as a push-pull amplifier. Class-C amplifiers are biased such that conducting current in the transistor flows for less than 50% of the incident input voltage signal. This allows for the highest efficiency at the expense of power gain and with the highest level of nonlinear operation. Class-C biased transistors are actually "off" without the presence of an RF signal on the input. For use in radar transmitter amplifiers, the Class-C amplifier offers higher efficiency over Class-A, -B, or -AB. In practice, they can be made to be "self-biasing" and have been the preferred class of operation for the silicon bipolar transistors used at UHF, L band, and S band. Because this class of operation is inherently nonlinear, as the transistor modulates between being off and saturated through each RF cycle, the harmonic content is high, and appropriate filtering of undesired higher order spectral content must be applied at the output of the transmitter. The amplifier Class-D, -E, -F, and -G are high efficiency switching amplifier configurations that require specialized termination of the signal harmonics (filtering) in order to maximize the amplifier efficiency. These can be complicated hardware implementations but may be warranted where incremental improvement in efficiency brings benefit to the transmitter system.

Silicon BJTs that operate in the HF through S-band frequency ranges are commonly biased either Class-B or Class-C. Class-C operation is the preferred mode because the RF output power of the amplifier is maximized for a given prime power input. In general, the base-emitter junction is reverse-biased, and collector current is drawn for less than half of an RF cycle. Collector current is drawn only when the input voltage exceeds the reverse bias across the input and the output voltage is developed across a resonant-tuned load. The net result is high amplifier efficiency. The practical implications of a Class-C-biased amplifier stage are as follows:

- No quiescent dc current is drawn while the device is not being driven, such as in the radar receive mode. Hence, there is no power dissipation in the amplifier while the transmitter is operating in this mode.
- Only one power supply voltage is necessary for the collector terminal of the transistor. The Class-C operation is a self-bias, wherein the transistor draws collector current only when the RF voltage swing on the input exceeds the built-in potential of the emitter-base junction. Additional reverse biasing may be introduced as a result of the voltage drop induced by current flow across parasitic resistance of the base or emitter-bias return, and in common base operation, this will result in degraded power gain.

- Class-C-biased amplifiers are very sensitive to any deviations from the nominal operating point. Class-C-biased amplifiers exhibit sensitivity to RF drive level and load impedance that may degrade the output pulse characteristics. A single-stage Class-C biased BJT amplifier will typically exhibit a very narrow "linear" transfer characteristic; the linear region may exist over only a narrow 1- to 3-dB window of RF input drive. This becomes strikingly critical when several Class-C-biased stages are cascaded in series, as is common in most amplifier configurations. The final tier of output transistors in a serial amplifier chain must be driven into saturation by the preceding stages, and the drive level must be held relatively constant as a function of time and temperature. Since these devices exhibit this narrow operating range, small decreases in the input RF drive level to a multistage amplifier may bring the final tier of devices out of saturation. Failure to control these conditions accurately can result in unacceptably degraded output pulse fidelity.

In a very simple sense, the design of an amplifier module consists of matching the power transistors to the proper impedance level and then combining the power levels at these impedances. A typical packaged power transistor has low input and output impedances that must be transformed up to higher level, usually 50 ohms. Thus, the typical amplifier design task must address both low-loss and inexpensive reactive impedance-transforming networks that can provide the proper source and load impedances to the transistor. The common medium for providing this function is a microstrip transmission line. Microstrip is a quasi-TEM mode transmission-line medium that requires photolithographically defined lines on a low-loss, high-quality dielectric substrate. Reactive components that are necessary as impedance-matching elements can be approximated in the microstrip format. An inexpensive reactive matching network can be formed by using an interconnected pattern of microstrip elements. Shunt- and series-connected inductive reactances as well as shunt capacitive reactances, are the most easily fabricated and most frequently used matching elements.

Power Combining. A power combiner coherently adds together the RF output voltages of individual amplifiers and delivers the sum total of the modules' output power, minus the losses of the combiner, to a single port. The outputs of identical single-stage power amplifiers are commonly summed by using well-documented[28] splitting and combining techniques (Figure 11.14). These techniques also address isolation between paralleled amplifiers. Having isolation between adjacent ports means that if one device fails, the power combiner provides a fixed load impedance to the remaining device; however, half the power of the remaining active device will be dissipated in the isolation resistor of the combiner. In order to achieve the most efficient combining of parallel amplifier stages, the phase and amplitude balance of individual stages should be as similar as possible. Any deviations from identical phase and amplitude balance result in power output lost to the resistive terminating port of the combiner. The power lost to similarity differences, from either phase or amplitude, is dictated by vector addition and is given by:

$$P_{\text{LOST}} = 20*\log\left(\left(\text{SQRT}\left(P_1^2 + P_2^2 + 2*P_1*P_2*\text{COS}(\theta)\right)\right)/(P1 + P2)\right) \quad (11.4)$$

where θ is the phase difference in degrees between two amplifiers that are summed together, SQRT indicates the "square root of,", and P_1 and P_2 are the power levels of each amplifier in watts. Figure 11.15 quantifies the impact of lost performance due to phase or amplitude imbalance.

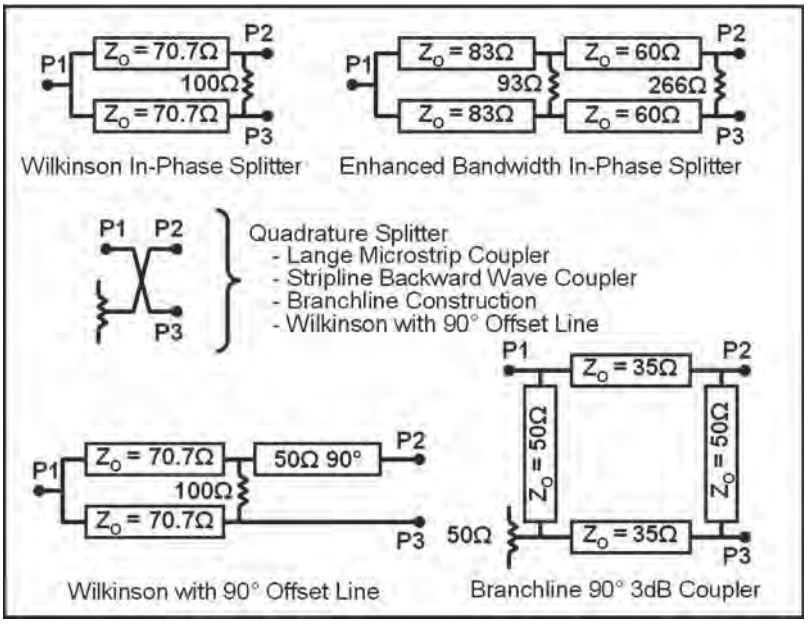


FIGURE 11.14 Common microwave power-combining circuit topologies that are used to provide isolation among adjacent parallel amplifiers in a corporate combining structure

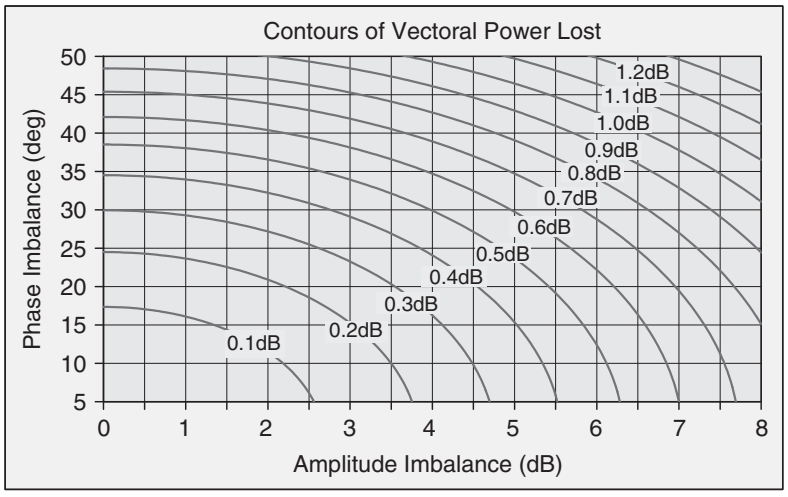


FIGURE 11.15 Contours of power lost to the isolation load resistor of an isolated power combiner for a range of amplitude and phase imbalances between two combined amplifiers. With an amplitude imbalance of 1 dB and a phase imbalance of 30°, approximately 0.31 dB of power will be lost to the isolating termination of the power combiner.

In general, the requirements for a power combiner are

- The combiner should have low insertion loss to maximize transmitter efficiency.
- The combiner should have RF isolation among ports, such that failed modules do not affect the load impedances or combining efficiency for the remaining functioning modules.
- The combiner should provide a controlled RF impedance to the amplifier modules, such that the amplifier characteristics are not degraded.
- The dissipated power capability of the power combiner terminations should be sufficient to accommodate any combination of power amplifier failures.
- The mechanical packaging of the power combiner should allow modules to be repaired easily. The packaging should also provide short, equal phase and low-loss interconnections between the amplifier modules and the combiner.

Power combiners may be either isolative or reactive designs. In isolative designs, any imbalance or difference between the phase and amplitude of the voltages that are being combined is directed to a resistive termination. The net result is that a constant load impedance is presented to the amplifier under all conditions even when an adjacent amplifier in a combining tier has failed. In a reactive combiner design, any imbalance in power or phase between two input signals results in reflected power and increased VSWR to the module driving it. Higher than desired frequency-dependent phase and amplitude ripple may result from improper use of this configuration.

A splitter and combiner network may also provide serial isolation among cascaded amplifier stages as well as parallel isolation. For example, when a Class-C biased transistor is pulsed, it passes through its cutoff, linear, and saturation regions. Consequently, the input and output impedances are dynamically varying, and the input impedance changes very dramatically. The dramatically changing input impedance will present an undesirable load to the preceding amplifier stage supplying the RF drive power. This may very well send the previous stage into unwanted oscillation. However, a quadrature splitter network, i.e., a power divider that provides a 3 dB split as well as a 90° phase offset, can be used to provide a constant impedance at the input to the splitter regardless of the individual amplifier input impedances (Figure 11.16). This ensures that a driver amplifier stage is presented with a well-matched load.

Typical RF transmission media that are used in the construction of high-power combiners include coaxial transmission lines, microstrip or stripline transmission lines, or waveguide. The choice of transmission medium is generally a function of many parameters, including peak and average power-handling capability, operating frequency and bandwidth, mechanical packaging constraints, and, of course, the overall loss that can be tolerated. More often than not, a combiner design utilizes a hierarchy of cascaded designs to sum the outputs of many modules; however, unique configurations that sum many ports to a single port have been built.

Amplitude and Phase Sensitivities. The phase and amplitude sensitivity of transistor amplifiers to power supply ripple may impact the MTI improvement factor that can be attained. In a multistage amplifier, the phase errors due to power supply sensitivity of serially cascaded stages will add. In addition, careful design must take into account interactions that can occur as a result of the many cascaded stages of solid-state amplification. These include the following:

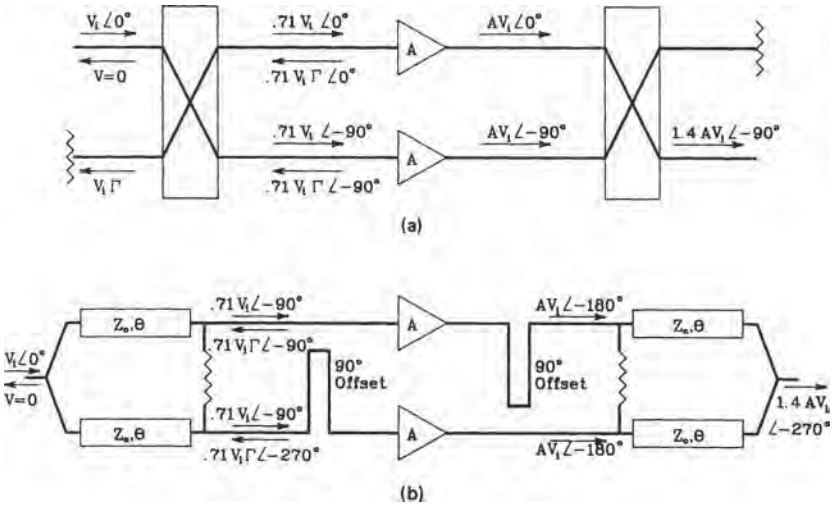


FIGURE 11.16 Power amplifier combining configurations that provide minimum input port reflected power: (*a*) quadrature-coupled amplifier pair and (*b*) split-T amplifier pair with a 90° offset. The amplifier input voltage reflection coefficient is given as Γ and the amplifier voltage gain as *A*. (*Reprinted with permission from E. D. Ostroff et al*., Solid-State Transmitters, *Norwood, MA: Artech House, 1985*.)

- Phase errors in cascaded stages simply add. However, it may also be possible to arrange them to cancel by proper phasing of power supply ripples for different stages. Similarly, in a stage with *N* modules in parallel, each with its own high-frequency power-conditioned power supply, the overall phase ripple can usually be assumed to be reduced by a factor equal to the square root of *N* if the power supply clocks are purposely not synchronized.
- Because of saturation effects, amplitude errors in cascaded stages do not simply add. However, amplitude errors in driving stages will cause drive-induced phase variations in the following stages, as noted above, all of which must be counted.
- Time jitter in cascaded stages simply adds unless the stages are arranged to cancel or to be root-sum-squared. In addition, amplitude fluctuations in the RF drive will also cause drive-induced jitter, which may even exceed power-supply-ripple-induced jitter, so this factor must be carefully measured.

Spectral Emissions. When a rectangular RF drive pulse is applied to a single module, the amplifier will typically show rise and fall times that are on the order of nanoseconds. The output signal spectrum of this pulse shape may not meet spectral emissions requirements, and it may be necessary to slow the rise and fall times. However, the amplifier operating region of optimum efficiency occurs as the transistor is driven into saturation, and for a large transmitter, there may be numerous tiers of cascaded saturated amplifiers. With so many cascaded saturating amplifiers, it becomes very difficult to control the rise and fall times as a result of the nonlinearity that is introduced into the power transfer function for the transmitter. Consequently, an input pulse shape with very exaggerated slow rise and fall times may be necessary to achieve the desired output-pulse spectral composition.

11.5 DESIGNING FOR THE SOLID-STATE PHASED ARRAY TRANSMITTER

In contrast to the design of the solid-state bottle transmitter where significant losses can accrue in the combining circuitry, the solid-state phased array antenna uses individual transmit/receive (T/R) modules with internal phase shift capability. Each T/R module is located behind an associated radiating element in a two-dimensional array. In this fashion, the beam is more efficiently formed in space and one avoids the losses than can accumulate in corporate combining. The transmit/receive (T/R) module, regardless of complexity, has five fundamental functions: (1) to provide gain and power output in the transmit mode, (2) to provide gain and low-noise figure in the receive mode, (3) to switch between transmit and receive states, (4) to provide phase shift for beam steering for both transmit and receive paths, and (5) to provide self-protection for the low-noise amplifier.

The first T/R module was developed by Texas Instruments in the mid-1960s as part of the Molecular Electronics for Radar Applications (MERA) program initiated by the U.S. Air Force to determine the feasibility of using X-band T/R modules in a solid-state phased array radar.[29] As a result of continuous development, phased arrays are used in multiple military and communications systems. The advantages of a phased array transmitter include (1) the ability to have multiple independently steered beams from a single aperture, (2) the speed of electronic versus mechanical beam locations, and (3) the efficiency of utilizing space combining instead of performing the power combining before the antenna. Block diagrams of representative T/R module functions are shown in Figure 11.17. Functionally, these are all equivalent, but the partitioning of circuit functions is dependent on the capability of the MMICs used, and different implementations may be required to address a key reliability requirement or a key performance parameter. For example, the use of a single high performance power amplifier may obviate the need for combining two lesser power amplifiers together to achieve the same performance. These represent cost, capability, and availability trades that might be exercised by the T/R module architect.

Microwave Monolithic Integrated Circuits (MMICs). During the 1990s, it was the reduction-to-practice of the microwave monolithic integrated circuit (MMIC) that enabled most high frequency phased arrays to be realized. MMIC use in T/R module design has enabled bold new module configurations, and hence phased array systems, to be envisioned. Because some of the more complex functions in the generic T/R module block diagram can be fabricated by using MMIC technology, the components that can be realized through the use of this technology can be employed to create system architectures that are difficult, if not impractical, to design with other, less integrated technologies. The MMIC design approach utilizes active and passive devices that have been manufactured by using a single process. Active and passive circuit elements are formed on a semi-insulating semiconductor substrate through various deposition schemes. The monolithic approach to circuit design inherently offers several advantages:

- **Low-cost circuitry** Component assembly is eliminated because complex circuit configurations using both active and passive components are batch processed on the same substrate.

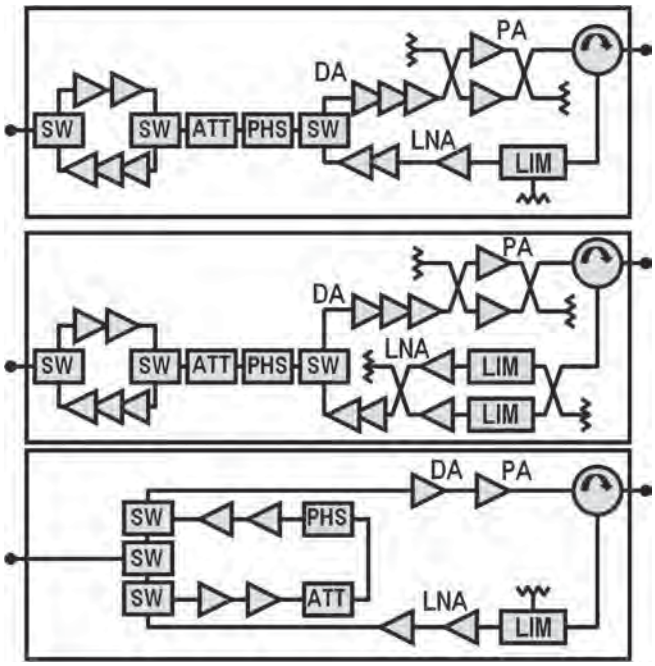


FIGURE 11.17 Common T/R module configurations make use of power amplifiers, low-noise amplifiers, duplexers, switches, and controls to enable on-face beam steering in a phased array antenna. Architecture variations may result from component capability differences as well as performance and packaging constraints.

- **Increased reliability** Batch-processed components lead to a reduced number of parts from the reliability standpoint and hence to increased operating lifetimes.
- **Increased reproducibility** Circuitry that is batch-processed or circuits that originate from the same wafer exhibit consistent electrical characteristics from component to component.
- **Small size and weight** Integration of active and passive components onto a single chip results in high-density circuitry with multiple functions on a single chip. Overall, the T/R module can be made much smaller than with discrete components.

The partitioning of T/R module circuit functions onto monolithic chips usually represents a tradeoff among several design issues, and the resultant circuit configurations represent a compromise among the goals of optimum RF performance, high levels of integration, and fabrication yields that are consistent with processing capabilities of GaAs MMICs. Among the single-chip circuit designs that have been reported from UHF through millimeter-wave frequencies are power amplifiers, low-noise amplifiers, wideband amplifiers, phase shifters, attenuators, T/R switches, and other special function designs. Noteworthy design considerations for these MMIC functions are described next.

Power Amplifiers. (1) The area consumed by combining in parallel the total number of gate fingers (i.e., total gate periphery) is usually at a premium. For high-power design, the load impedance presented to the final device must be carefully chosen such that power output and efficiency are maximized. Also, too much gate periphery may increase the chip area such that cost of the component becomes unattractive. (2) Losses in the output circuit of the final stage can significantly reduce power output and efficiency. Off-chip matching may be necessary to maximize power output for a given design. (3) GaAs is a poor thermal conductor. Power FET design that addresses thermal management is required. Adequate heat sinking of the chip is mandatory and may become a limiting factor in high performance designs. (4) Careful attention must be paid to unplanned voltage stresses on the power amplifier, either from transient induced effects or load impedance variations in order to maintain the desired reliability. (5) For efficient multiple-stage designs, it is necessary that the final stage of the amplifier reach saturation before the preceding stages. This must be addressed in the circuit design.

Low-Noise Amplfiiers. (1) Multiple stage linear designs require proper device sizing of successive stages in order to maintain low intermodulation distortion products. (2) Circuit losses on the input, before the first stage, degrade the noise figure of the design; therefore, some designs utilize off-chip matching. (3) The best noise-figure usually requires a bias condition that is closer to the pinch-off voltage of the FET than for a power amplifier. The pinch-off voltage is the voltage that when applied to the gate terminal causes the current in the transistor channel to stop flowing. Thus, the transistor is "pinched-off" and variability around this operating point can cause large circuit performance variability if designed poorly. Both gain and noise figures are highly dependent on the pinch-off voltage when the FET is biased close to pinch-off. Because the pinch-off voltage can vary among devices from the same wafer, the bias condition must be chosen carefully. Gain and noise figures are usually traded off against repeatable performance. Examples of an L band two-stage low-noise amplifier GaAs MMIC and an X band power amplifier GaAs MMIC are shown in Figure 11.18 and Figure. 11.19, respectively.

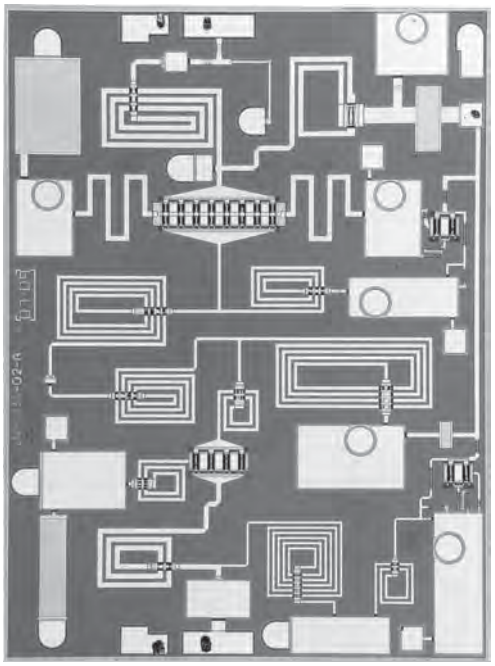


FIGURE 11.18 L-band low-noise amplifier MMIC. Shown here are the spiral inductors, metal-nitride-metal capacitors, and via-hole connections to ground. (*Photograph courtesy of Raytheon Company*)

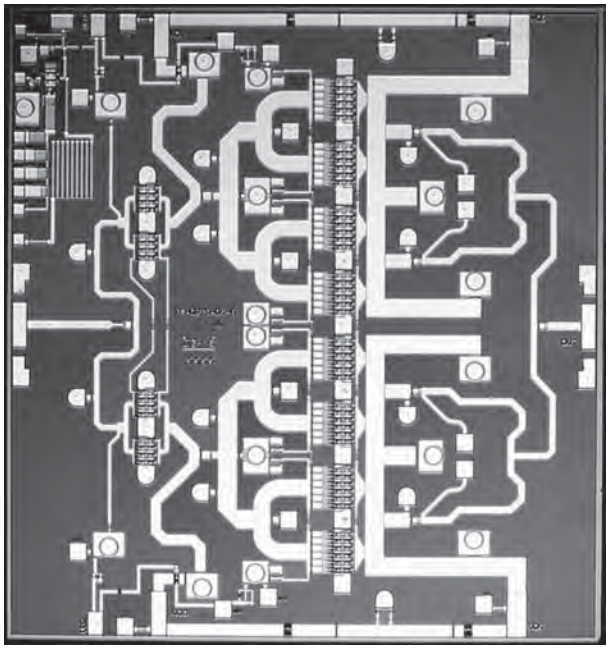


FIGURE 11.19 X-band two-stage power amplifier MMIC showing parallel combination of FET cells in the output stage. This example was fabricated on 100 μm- thick GaAs substrate. The RF lines are microstrip-format TEM-mode transmission lines. (*Photograph courtesy of Raytheon Company*)

Transmit/Receive Switching. (1) For switching applications, the FET design should be chosen such that the ratio of OFF-ON resistance of the FET is kept as high as possible. The channel length largely determines the ON resistance and hence the insertion loss of the device. The tradeoff between short gate length (thus lower processing yield) and insertion loss must be examined. (2) The value of the parasitic drain-source capacitance will affect the OFF-state isolation of the device. This capacitance depends largely on the source-drain spacing of the FET geometry. Critical applications are usually only the front-end switching configurations in a T/R module, i.e., before the receive low-noise amplifier or after the transmit amplifier.

Phase Shifters. (1) Digitally controlled phase-shifter designs generally utilize either a switched-line or a loaded-line circuit configuration, using either distributed transmission-line components or lumped-element equivalent circuits, to achieve multiple-bit phase shifting. Switched-line configurations rely on FET switches to switch lengths of transmission line in and out of the circuit and are typically used for higher frequencies where less chip area is needed. Loaded-line configurations use the switched FET parasitics as circuit elements to introduce the necessary phase changes.

The typical processing and construction sequence for a MMIC chip is fairly similar among the GaAs foundries (Figure 11.20). In this figure, the active channel region

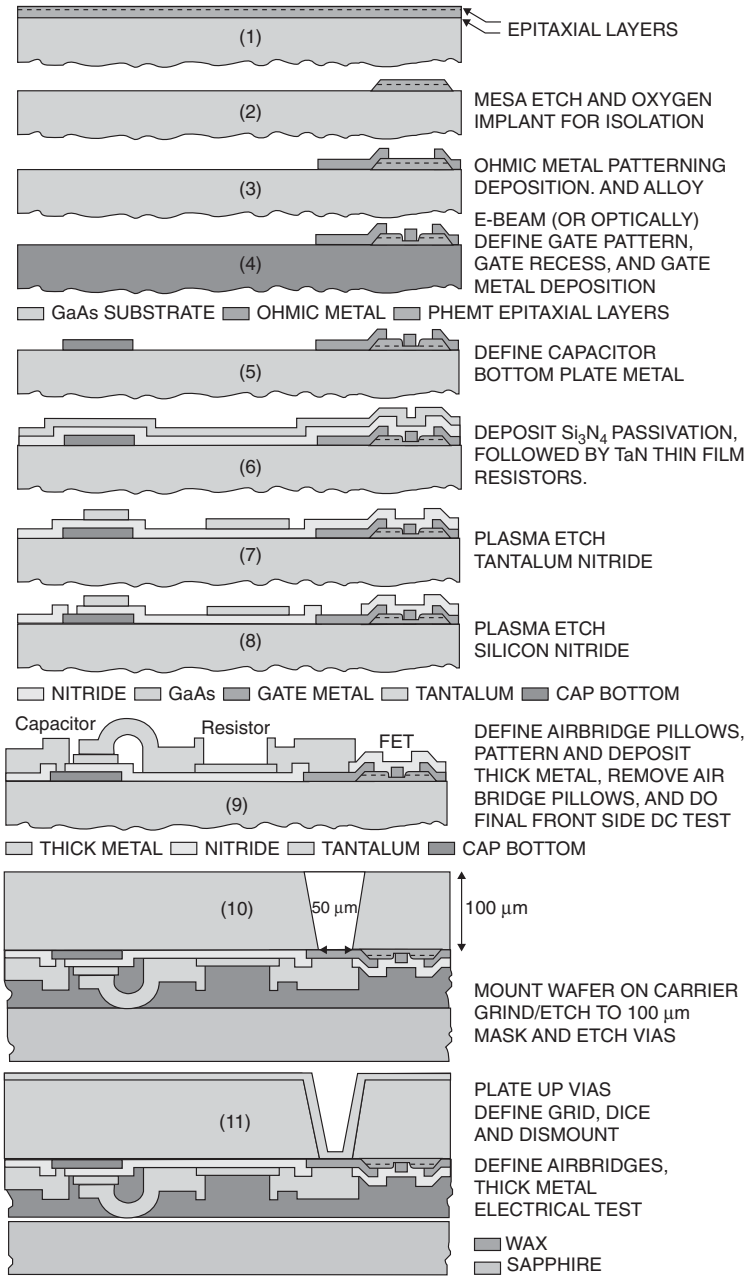


FIGURE 11.20 GaAs MMIC processing makes use of deposition and etching techniques to first define the active channel region of the transistor (1–4), followed by the deposition of metals, dielectric layers, and resistive layers forming the passive components (5–8). Then, thick metal interconnects (9) are introduced, followed by backside processing to connect RF ground to the topside components (10–11).

of a FET is delineated by any of several patterning techniques on a semi-insulating GaAs substrate such as ion implantation or molecular beam epitaxy. Once the FET has been defined, a combination of deposited dielectric films and metal layers is used to form the passive components (such as metal-insulator-metal capacitors) and also to interconnect all the elements of the circuit. Standard libraries of circuit elements may include FETs (used as linear amplifiers, low-noise amplifiers, saturating power amplifiers, or switches), resistors, capacitors, inductors, diodes, transmission lines, interconnects, and plated ground vias.

Transmit/Receive Module Characteristics. The impact of antenna array electrical requirements on the packaging of MMIC components into a T/R module is fundamental. The periodic nature of digital phase shifting at each radiating element can create multiple distinct locations in space where parasitic beams (grating lobes) can occur. In array design, this is avoided if the radiating element spacing (d) is less than that described by

$$d < (\lambda^*(1 + \sin(\theta))^{-1} \tag{11.5}$$

Where d is the spacing between adjacent radiating elements, λ is the wavelength of the highest operating frequency, and θ is the maximum scan angle of the array. For hemispherical phased array coverage, the maximum scan angle can be upwards of $\pm 60°$, depending on the number of array faces used in the system configuration. Thus, for an X-band array that requires scanning to large angles, the spacing between radiating elements, and by implication, the maximum spacing available for T/R modules when they are aligned behind the radiating elements must be on the order of 0.5 inches or less. Alleviations in packaging may be allowable if the scan volume is not required to extend to a full field of view. Values of element spacing that satisfy Eq. 11.5 are shown as a function of scan angle for some of the common radar frequency bands through mm-wave frequencies in Figure 11.21. The implication of this graph is that full T/R module functionality must be packaged into the space and volume behind the planar array, and this requirement can pose very difficult challenges to the T/R module

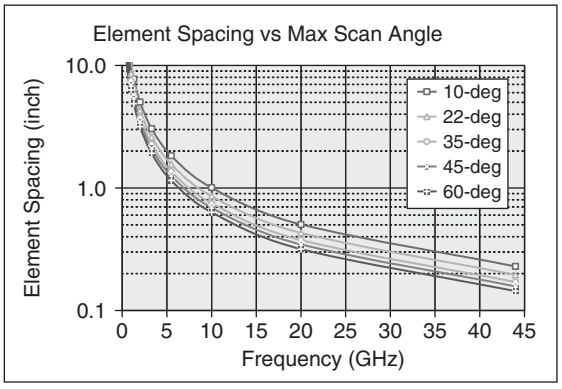


FIGURE 11.21 Maximum operating frequency and worst-case scan angle define the maximum allowable distance among adjacent radiating elements; T/R modules fitting behind each element are constrained by these spacings.

designer in order to satisfy the RF electrical, dc electrical, thermal, and reliability requirements. Packaging of MMIC components into the T/R module must take into consideration[30,31,32] multiple elements as they impact the electrical performance.

Power Conditioning Considerations. Pulsed transmit amplifiers can consume very high dc currents and special design attention must be paid to parasitic inductance that can generate very high voltage spikes and cause damage to MMIC power amplifiers. In addition, the dc power source must include appropriate energy storage, sometimes locally in the module, in order to support the minimum voltage pulse droop as a function of time.

Environmental Protection Considerations. MMIC components utilize thin-film metal deposition techniques to delineate the very fine features that make up the microwave circuitry. These features are susceptible to potential short-term failure due to corrosion, metal migration, and dendritic growth if there are voltages on the circuitry when exposed to an atmosphere that causes moisture to condense on the circuitry. Thus, a hermetic package with a dry, nitrogen-filled interior is usually employed to ensure long-term reliability. Hermetic packaging also brings with it the undesirable effect of trapping inside the housing any molecular content that outgases into the interior cavity. In particular, hydrogen can be present in the interior metal plating and has been known to cause long-term reliability concerns in some GaAs amplifiers. One solution involves the use of an internal hydrogen getter to counterbalance the reliability impact. A getter is a material included in the module housing to absorb residual hydrogen.

Mechanical Packaging Considerations. The T/R module housing must be made of materials that provide for adequate thermal management and long-term reliability. Materials that simultaneously support exposure to shock, vibration, temperature cycling, and adequate thermal management must be used. Materials that match very closely the coefficient of thermal expansion (CTE) of the semiconductor material must be used in the design of the housing such that cracking of the semiconductor devices does not occur during thermal cycling that happens during normal operation or during temperature changes during assembly and test.

Electrical Interconnection Considerations. The interconnection of MMIC chips within the T/R module must utilize controlled impedance transmission lines with low insertion loss. Thus, some combination of high-quality microwave dielectric material must be integral to the microwave electrical and mechanical design of the module. Attention to the coefficient of thermal expansion and manufacturability issues will impact the choice of usable materials. T/R modules also generally require as many as 6–12 control or bias connections in order to interface with amplifiers, control circuitry, and phase shifters. The interconnection density, especially at higher frequencies, can become a packaging design challenge. At frequencies above 20 GHz, the use of conventional connectors is usually prohibitive due to the small width available in full field-of-view arrays.

Manufacturability Considerations By definition, the use of MMIC components invokes a microelectronic assembly, test, and handling manufacturing infrastructure. The manufacturing of low-cost T/R modules is paramount to being able to effectively produce affordable arrays. Design methodologies, such as statistical performance

representation, are often exploited to maximize functional yield. The integration of design and manufacturing processes is a key component to successful execution of product manufacturing.

11.6 SOLID-STATE SYSTEM EXAMPLES

PAVE PAWS (UHF Early Warning Radar). The PAVE PAWS (AN/FPS-115) system is a UHF solid-state active aperture phased array radar that was built for the Electronic Systems Division of the U.S. Air Force by the Equipment Division of the Raytheon Company during the late 1970s.[33] The radar is a long-range system with a primary mission to detect and track sea-launched ballistic missiles. The two-faced radar uses 1792 active T/R modules per face, and each module interfaces with a dipole antenna element. Extra elements and a narrow beam are used on receive, and upgrade capability has been included for the future installation of up to 5354 T/R modules per array face. The peak power output from each face, when populated with 1792 modules is 600 kW, and the average power output is 150 kW.

Among the 1792 modules per face, groups of 32 T/R modules are operated as a subarray. In transmit, a high-power array predriver is used to drive 56 subarray driver amplifiers. Each of these power amplifiers provides enough RF drive for all 32 modules in one subarray. In receive, the signal from each of the 56 subarrays is fed into a receive beamforming network.

The T/R module contains predriver, driver, and final transmit amplifiers, transmit/receive switching, low-noise amplifiers, limiter, phase shifters, and logic control. The T/R module block diagram is shown in Figure 11.22, and a photograph is shown in

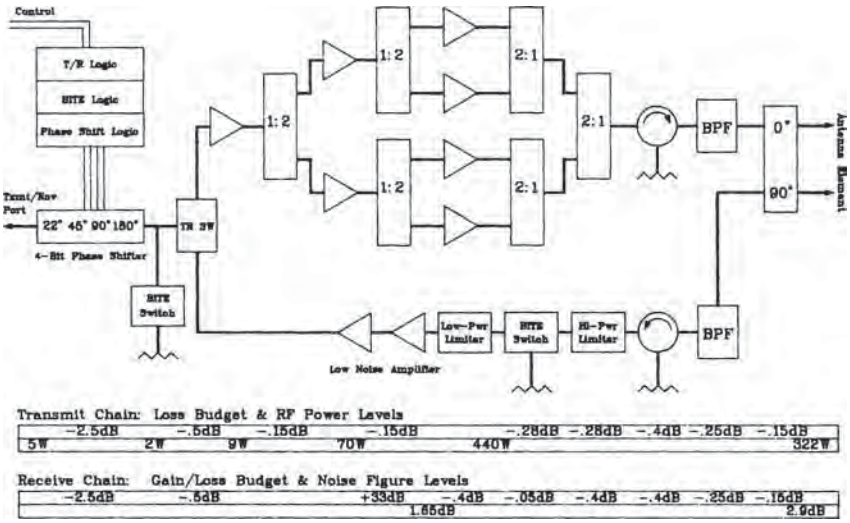


FIGURE 11.22 Block diagram of the PAVE PAWS transmit/receive module shows a 1-2-4 transistor driving configuration for the transmit amplifier and a quadrature splitter on the output to generate a polarized feed to the radiating element.

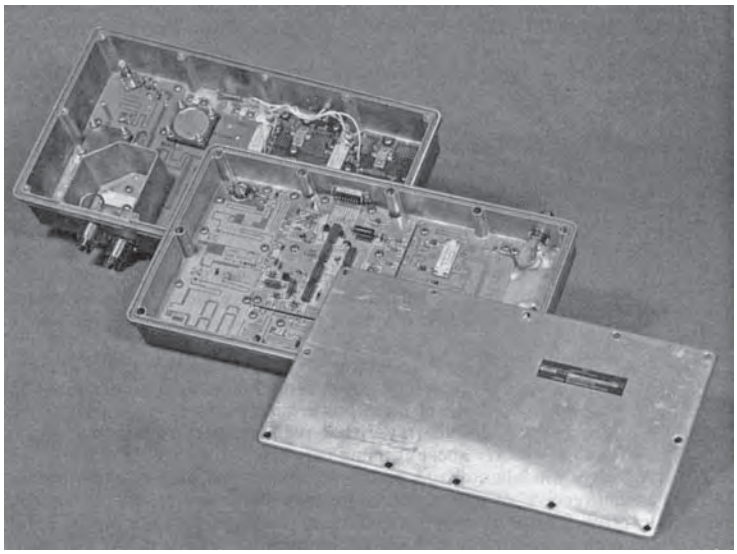


FIGURE 11.23 PAVE PAWSUHF T/R module consists of transmit module and receive module in a nested configuration of cast aluminum housings (*Photograph courtesy of Raytheon Company*)

Figure 11.23. The transmitter portion of the T/R module contains seven silicon bipolar power transistors, operated Class-C from a 31 V dc power supply. The transmit amplifier chain consists of a predriver transistor feeding two driver transistors, in turn feeding four final transistors, i.e., a 1-2-4 configuration. Each of the four final stages delivers 110 W peak for 16-ms pulse widths at duty cycles up to 25%. More than 180,000 transistors have been built into more than 25,000 modules. Future upgrades of this design in the BMEWS arrays will use the more powerful and efficient Si LDMOS FET technology for enhanced performance capabilities.

AN/SPS-40 Shipboard Search Radar. The AN/SPS-40 was an existing UHF, tube-type, long-range, 2D shipboard search radar system, for which a new solid-state transmitter was built during the 1980s to replace the tube. The solid-state transmitter was built for the Naval Sea Systems Command by the (then) Westinghouse Electric Corporation.[34] The existing waveform from the original transmitter was not changed, and the solid-state unit was installed as a direct retrofit. This was not quite as difficult as usual, because the tube-type system already used long pulses and pulse compression with a duty cycle of nearly 2%, which is much higher than older 0.1% duty cycle systems. Although it may have been desirable to go to a higher duty cycle and lower peak power to make the solid-state retrofit easier, the Navy preferred not to have to modify the rest of the system.

The 250-kW peak power transmitter used a total of 128 high-power amplifier modules, which, along with power combining, predrivers, drivers, and control circuitry, were housed in three separate cabinets. There were 112 final power output modules arranged in two groups of 56. Each module (Figure 11.24) produced 2500 W peak and 50 W average for a 60-μs pulse width at a 2% duty cycle. Drive power for the two

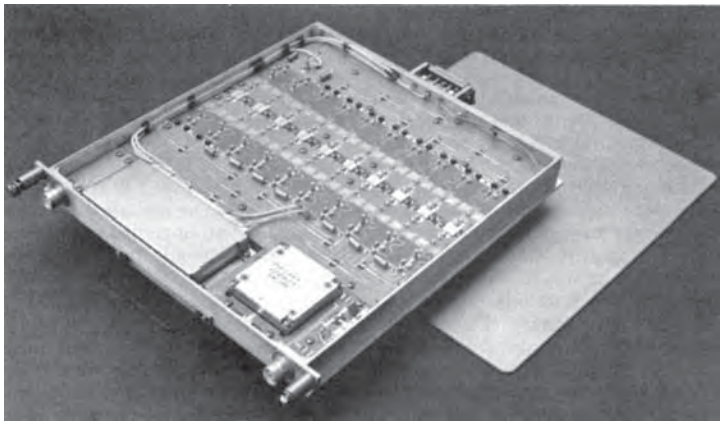


FIGURE 11.24 AN/SPS-40 transmitter amplifier module (*Photograph courtesy of Westinghouse Electric Corporation*)

banks of final output modules, 17.5 kW, was provided from the combined outputs of 12 more identical modules in the driver group. Predrivers and a redundant preamplifier were used as preceding drive stages.

The power amplifier module consisted of ten identical silicon bipolar power transistors arranged in a 2-driving-8 amplifier configuration to develop more than 2500 W peak power output over the 400 to 450 MHz frequency bandwidth. Each transistor was a 400 W peak-power device that was operated in a balanced push-pull circuit design. By using a push-pull configuration, the circuit designers alleviated some of the low impedance-matching problems normally associated with very high power transistors. The RF input drive to the module was 120 W peak and was used to drive two devices. A combined power level of greater than 600 W was split eight ways to drive the eight identical output stages. Losses in the output circulator, final power combining, and the fault detection circuitry reduced the combined power level to 2500 W. Output modules were liquid-cooled for normal operation, but an emergency backup forced-air cooling was provided in the event of a primary-cooling-system failure. The dissipated heat could be tolerated because the system operated at a low duty cycle.

The power combining for each output cabinet consisted of 56:1 combiners. The reactive power combiner consisted of seven groups of 8:1 combiners fabricated in air stripline using 0.5-in ground-plane spacing. The seven outputs were combined by using a single 7:1 air stripline combiner with 1.0-in ground-plane spacing. The 130 kW outputs of the two 56:1 combiners were combined in a single 2:1 isolated hybrid that was manufactured by using a coaxial transmission line. The advertised losses of the 2:1 and 56:1 combiners were 0.1 dB and 0.25 dB, respectively.

RAMP (L-Band Air Traffic Control Transmitter). The Radar Modernization Project (RAMP) radar system is an L band system built by the Raytheon Company during the late 1980s to replace the earlier primary and secondary surveillance radars used for air traffic control by Canada's Ministry of Transport.[35,36] The primary surveillance radar consists of a rotating reflector, horn-fed by a solid-state transmitter, and interfacing with redundant receive channels with receiver-exciters and signal processors. The primary surveillance radar operates between 1250 and 1350 MHz with a 25-kW peak

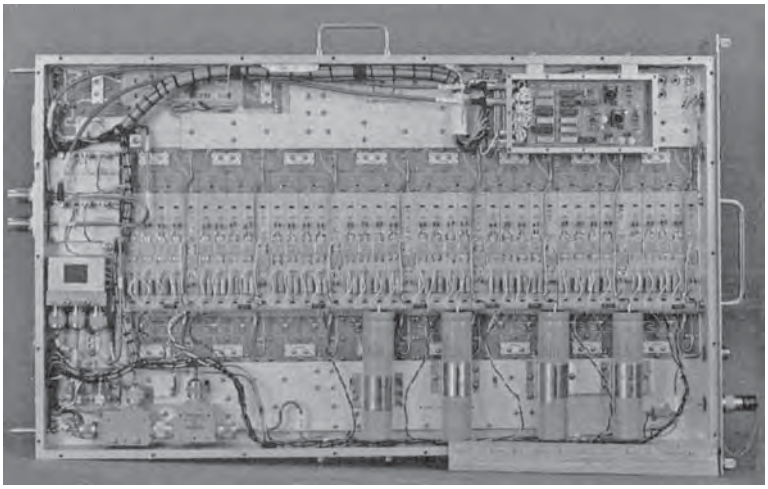


FIGURE 11.25 RAMP transmitter amplifier module (*Photograph courtesy of Raytheon Company*)

power output and provides radar coverage to 80 nmi and to an altitude of 23,000 ft with an 80% probability of detection for a $2m^2$ target; with azimuth and range resolution to 2.25° and 0.25 nmi, respectively. The receiver-exciter efficiently utilizes the transmitter solid-state devices with a high duty-cycle waveform. A pair of pulses is used in the frequency-agile system, and target returns are processed by a moving-target detector. The pulse pair consists of a 1-μs single-tone pulse that provides coverage to 8 nmi and a 100-μs nonlinear chirp pulse that provides coverage to 80 nmi. The 100-μs pulse is compressed to 1 μs such that high duty cycle is achieved without compromising range resolution. The transmitter consists of 14 modules, each capable of 2000 W power output (Figure 11.25), that are combined to produce the greater than 25 kW peak power level. Two modules and a 33 V dc power supply make up a single transmitting group. The module consists of a 2-driving-8-driving-32 transistor (2-8-32) amplifier configuration of silicon bipolar power devices. The two final output devices and the eight driver devices are 100 W transistors capable of operating up to a 10% duty cycle over the 100-MHz bandwidth at collector efficiencies greater than 52%. Each module is air-cooled, and the measured efficiency is greater than 25% when the module is operating at an 8.2% average duty cycle. Module power gain is greater than 16 dB. A circulator is used on the output port to protect the 100 W devices from antenna-generated reflections, and control circuitry has been included to switch off modules in the event of cooling-system failure. A 14:1 high-power replicated combiner built by using a combination of reactive and resistive power-combining techniques in air-dielectric stripline, is employed to sum the module outputs to the 25 kW level.

REFERENCES

1. M. Meth, "Industrial assessment of the microwave power tube industry," Department of Defense Report, April 1997, p. 3.
2. V. Granatstein, R. Parker, and C. Armstron, "Scanning the technology: Vacuum electronics at the dawn of the 21st century," *Proceedings of the IEEE*, vol. 87, no. 5, pp. 702–716, May 1999.

3. V. Gregers-Hansen, "Radar systems trade-offs, vacuum electronics vs. solid-state," in 5th International Vacuum Electronics Conference, April 27–29, 2004, pp. 12–13.
4. R. Symons, "Modern microwave power sources," *IEEE AESS Systems Magazine*, pp. 19–26, January 2002.
5. T. Sertic, "Idiosyncrasies of TWT amplifiers," presented at The Future of Electronic Devices Conference, Institute of Physics, March 22, 2003.
6. M. Hanczor and M. Kumar, "12-kW S-band solid-state transmitter for modern radar SYSTEMS," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, no. 12, pp 2237–2242, December 1993.
7. D. Rutledge, N. Cheng, R. York, R. Weikle, and M. DeLisio, "Failures in power combining arrays," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 7, pp. 1077–1082, July 1999.
8. L. B. Walker, *High Power GaAs FET Amplifiers*, Norwood, MA: Artech House, 1993, p. 5.
9. Hewlett-Packard Application Notes, High Frequency Transistor Primer, Part 3, Thermal Properties, p. 6.
10. H. Cooke, "Microwave transistors: theory and design," *Proceedings of the IEEE*, vol. 59, pp. 1163–1181, August 1971.
11. Vendor transistor datasheet, Integra Technologies, Inc, www.integratech.com.
12. Vendor transistor datasheet, Tyco Electronics, M/A-COM, www.macom.com.
13. Vendor transistor datasheet, STMicroelectronic, www.st.com.
14. Vendor transistor datasheet, Philips, www.datasheetcatalog.com.
15. N. Sakura, K. Matsunage, K. Ishikura, I. Takenake, K. Asano, N. Iwata, M. Kanamori, and M. Kuzuhara, "100W L-band GaAs power FP-HFET operated at 30V," in *IEEE Microwave Theory and Techniques Symposium Digest*, 2000, pp. 1715–1718.
16. T. Winslow, "Power dependent input impedance of field plate MESFETs," *Compound Semiconductor Integrated Circuit Digest*, pp. 240–243, 2005.
17. J. Huang, G. Jackson, S. Shanfield, A. Platzker, P. Saledas, and C. Weichert, "An AlGaAs/InGaAs pseudomorphic high electron mobility transistor with improved breakdown voltage for X- and Ku-band power applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 41, no. 5, pp. 752–758, May 1993.
18. K. Alavi, S. Ogut, P. Lymna, and M. Borkowski, "A highly uniform and high throughput double selective PHEMT process using an all wet etch chemistry," presented at GaAs MaTech Conference, 2002.
19. C. Snowden, "Recent development in compound semiconductor microwave power transistor technology," *IEE Proc-Circuits Devices Syst*., vol. 151, no. 3, pp. 259–264, June 2004.
20. D. Miller and M. Drinkwine, "High voltage microwave devices: An overview," presented at International Conference on Compound Semiconductor Mfg., 2003.
21. J. Zolper, "Scanning the special issue, special issue on wide bandgap semiconductor devices," *Proceedings of the IEEE*, vol. 90, no. 6, pp. 939–941, June 2002.
22. U. Mishra, P. Parikh, and Y. Wu, "AlGaN/GaN HEMTs—An overview of device operation and applications," *Proceedings of the IEEE*, vol. 90, no. 6, pp. 1022–1031, June 2002.
23. R. Trew, "SiC and GaN transistors—Is there one winner for microwave power applications?" *Proceedings of the IEEE*, vol. 90, no. 6, pp. 1032–1047, June 2002.
24. S. Allen, R. Sadler, T. Alcorn, J. Palmour, and C. Carter, "Silicon carbide MESFETs for high power S-band applications," in *IEEE MTT-S International Microwave Symposium*, June 1997, pp. 57–60.
25. Thomas Kazior (personal communication), Raytheon RF Components, August 2006.
26. Y. Wu and P. Parikh, "High-power GaN HEMTs battle for vacuum-tube territory," *Compound Semiconductor Magazine*, January/February 2006.
27. Colin Whelan (personal communication), Raytheon RF Components, August 2006.
28. H. Howe, *Stripline Circuit Design*, Norwood, MA: Artech House, 1974, pp. 77–180.
29. D. McQuiddy, R. Gassner, P. Hull, P., J. Mason, and J. Bedinger, "Transmit/receive module technology for X-band active array radar," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 308–341, March 1991.

30. G. Jerinic and M. Borkowski, "Microwave module packaging," in *IEEE Microwave Theory and Techniques Symposium Digest*, 1992, pp. 1503–1506.
31. B. Kopp, M. Borkowski, and G. Jerinic, "Transmit/receive modules," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 827–834, March 2002.
32. B. Kopp, C. Moore, and R. Coffman, "Transmit/receive module packaging: Electrical design issues," *Johns Hopkins APL Technical Digest*, vol. 20, no. 1, pp. 70–80, 1999.
33. D. Hoft, "Solid-state transmit/receive module for the PAVE PAWS phased array radar," *Microwave Journal*, pp. 33–35, October 1978.
34. K. Lee, C. Corson, and G. Mols, "A 250 kW solid-state AN/SPS-40 radar transmitter," *Microwave Journal*, vol. 26, pp. 93–96, July 1983.
35. J. Dyck and H. Ward, "RAMP's new primary surveillance radar," *Microwave Journal*, p. 105, December 1984.
36. H. Ward, "The RAMP PSR, a solid-state surveillance radar," presented at IEE International Radar Conference, London, October 1987.

Chapter 12
Reflector Antennas

Michael E. Cooley and Daniel Davis

Electronic Systems
Northrop Grumman Corporation

12.1 INTRODUCTION

Role of the Radar Reflector Antenna. Radar reflector antennas provide the means by which the transmit (receive) energy and its associated waveform is radiated into (coupled from) free space. In transmit mode, the antenna launches a guided wave from the transmitter into free space and typically focuses this radiated energy over a limited angular range or beamwidth. In receive mode, the reflector antenna operates in a reciprocal manner, receiving reflected radar target energy, i.e., echoes, from a limited angular range. These received echoes are then converted into guided waves that are amplified and subsequently processed in the radar receiver.

Typically, the radar reflector antenna must be designed to enable beam scanning over the field-of-view (FOV) via either mechanical or electronic means (or some combination of both). In Section 12.4, methods of electronic beam scanning (limited FOV) using array feeds are discussed. Thus, the radar reflector antenna performs several important functions: (1) It converts the guided wave from the transmitter to a radiated wave (or vice versa on receive); (2) It concentrates or collimates the radiated energy into a directive beam of specified gain and beamwidth; (3) It collects the reflected energy scattered from the radar target; and (4) It supports beam scanning via either electronic or mechanical means (or both).

Antenna Beam Scanning. For most radar applications, the trade or choice between a reflector antenna and a direct radiating phased array is typically driven by factors relating to scan rate, scan volume, and cost. Reflector antennas are typically employed in a radar when (1) slower scan rates are sufficient and mechanical scanning suffices, and/or (2) a very high gain (electrically large) aperture is required and a phased array, i.e., an electronic scanning array (ESA), is cost prohibitive, and/or (3) the required scan volume is limited and can be satisfied via use of an array-fed reflector. During the 1980s and 1990s, phase shifter and T/R module technology greatly matured and ESA costs dropped dramatically. These advances have resulted in increased interest and utilization of ESAs for wide scan radar applications, and array-fed reflectors where limited electronic scan suffices.

Advantages and Applications of the Radar Reflector Antenna. In the previous paragraph, the proliferation of ESA antennas in modern radar systems is linked to the

dramatic T/R module cost reductions and technology improvements. The improved performance of ESA radars is cited as a reason for decreased utilization of reflector antennas in many of today's radar system designs.

However, there are still applications where the reflector antenna is well suited to radar applications and will continue to find applications in the future. Three relevant examples of radar applications well suited to the use of reflector antennas are briefly described below.

Low Cost Radar. For very cost-constrained applications where mechanical scan rates suffice, reflector antennas are still the dominant choice. One such niche is commercial weather radar, e.g., NEXRAD and TDWR.

Very High-Gain, Long-Range Radar. For very high-gain radar applications, the cost of an ESA is typically still prohibitive, and the reflector provides an economical means of realizing such high gains. Two examples of long-range radar applications generally requiring very high antenna gains are (1) missile defense radar and (2) space-based radar.

Limited Scan Radars. Some radars operate over a limited field-of-view and/or the requirements dictate fast electronic scanning over a small FOV and slower mechanical scanning over a larger field of view. ESA-fed reflector architectures are well suited for such applications and are described in greater detail in Section 12.3. Three relevant examples are (1) missile defense radar, (2) space-based radar, and (3) ground-based search and track radar (1D azimuth electronic scanning suffices for some of these applications).

Classification of Reflector Antennas. Radar reflector antennas can be classified in various ways. One useful classification criteria is electrical design, i.e., the reflector optics configuration. Table 12.1 provides a summary level comparison of some common radar

TABLE 12.1 Comparison of Key Features of Reflector Architectures

	Single Reflector Parabolic	Cylindrical Reflector	Dual Reflector (Cassegrain or Gregorian)	Confocal Paraboloids	Spherical and Torus
Electronic Scanning (Escan)	• Limited in both Azimuth and Elevation • Achieved via Feed switching	• Wide 1D scanning • Uses ESA line source for wide 1D scanning	• Limited in both Azimuth and Elevation • Achieved via feed switching	• Uses planar ESA source • Typically limited, but tradable by varying magnification	• Potential for very wide 1D scanning (torus) or 2D Scanning (spherical) • Achieved via feed switching
Aperture Efficiency Feed Type(s)	• Medium to high • No escan: Single horn • Escan: Array (switched feeds)	• Medium to high • 1D ESA line source	• High • No escan: Single horn • Escan: Array (switched feeds)	• High • 2D ESA planar source	• Modest to low • Switched beam array on circular arc (torus reflector) or spherical arc (spherical reflector)
Blockage Concerns	• Mitigate feed blockage via offset geometry	• Mitigate feed blockage via offset geometry	• Mitigate feed blockage via offset geometry • Can move feed behind reflector	• Mitigate feed blockage via offset geometry	• Serious concern for very wide scan configurations

reflector architectures from this perspective. More detailed discussions of the characteristics of these architectures are included in Section 12.3 wherein each of these architectures is afforded a dedicated subsection. Another means of classification is via platform (vehicle) or site (ground-based, ship-based, airborne, or space-borne). The platform frequently drives mechanical and environmental requirements and often either enables or constrains the reflector size. In Section 12.3, reflector architectures are discussed and compared, and in Section 12.6, mechanical and environmental design considerations are addressed.

Chapter Synopsis. The balance of this chapter is divided into five sections. Section 12.2 summarizes the basic design principles and parameters governing reflector antenna design. Section 12.3 provides a brief overview of conic sections and classes of reflector systems and associated optics. Section 12.4 discusses various types of reflector feeds and related design principles. Section 12.5 describes reflector analysis and synthesis methods and associated design software packages. Finally, Section 12.6 briefly reviews mechanical design issues and considerations.

12.2 BASIC PRINCIPLES AND PARAMETERS

Fundamentally, reflectors are antennas that work on optical principles, on receive, focusing energy to a focal point as a lens does for light. On transmit, power emanating spherically from a low gain, broad-patterned feed is reflected and energy is collimated to form a plane wave thereby providing increased antenna gain and a narrower beamwidth. In discussing antennas, one can use either transmit or receive arguments because antennas are reciprocal devices. This means that both the transmit and receive performance of a passive antenna, e.g., patterns, gain, losses, etc., can be predicted using basic aperture antenna principles. In this section, these basic reflector design principles are reviewed via use of a canonical example.

Consider a parabolic-shaped reflector forming a circular aperture, fed by a horn at a central focal point. This simple reflector configuration has a surface shape defined by the equation

$$z = \frac{x^2 + y^2}{4f} - f \tag{12.1}$$

where f is the focal length and the vertex is located at $z = -f$. The resultant antenna beam pattern points in the positive z-direction. For a round reflector of diameter D, the edge or rim is a circle and the edge dimensions are on this circle defined by

$$x_{\text{edge}}^2 + y_{\text{edge}}^2 = \frac{D^2}{4} \tag{12.2}$$

This reflector is shown in Figure 12.1.

One frequently is interested in the angle α from the z-axis to points on the reflector surface:

$$\alpha = 2 \arctan\left(\frac{\sqrt{x^2 + y^2}}{2f}\right) \tag{12.3}$$

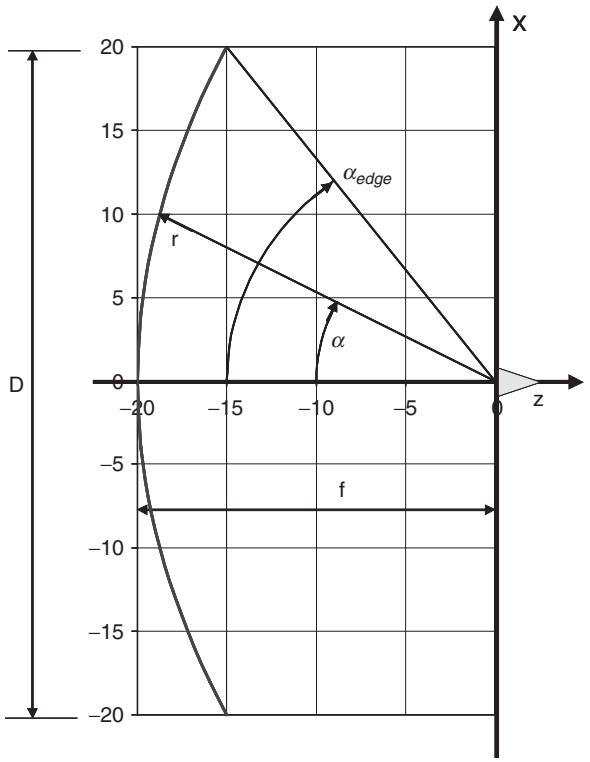


FIGURE 12.1 Parabolic reflector in *x*-*z* plane

Thus, the edge angle for a round reflector is

$$\alpha_{\text{edge}} = 2 \arctan\left(\frac{D}{4f}\right) \tag{12.4}$$

Another useful parameter is the distance r from the focal point to a point on the reflector:

$$r = f + \frac{x^2 + y^2}{4f} \tag{12.5}$$

Aperture Gain and Losses. The gain of the reflector antenna is one of its most important parameters. It is convenient to describe the reflector antenna gain with reference to the fundamental gain limit an aperture of area A. This limit, applicable to aperture antennas of sufficient electrical size (approximately 25 square wavelengths or greater), is the so-called aperture gain G_{ap},

$$G_{\text{ap}} = \frac{4\pi A}{\lambda^2} \tag{12.6}$$

where λ is the wavelength. For a round reflector of diameter D, this aperture gain limit is

$$G_{ap} = \left(\frac{\pi D}{\lambda}\right)^2 \tag{12.7}$$

In practice, it is sometimes useful to describe reflector gain by decrementing the aperture gain by subtracting various aperture radiation losses such as spillover, taper efficiency, feed blockage, reflector leakage, surface distortion, strut blockage, feed alignment, etc., from G_{ap}. These loss factors are described in detail later within this section.

Directive Gain and Feed Losses. Directive gain or directivity is a measure of the peak power relative to the average power radiated by an isotropic radiator, i.e., a radiator that radiates energy equally in all directions. Directive gain considers only the radiated power; thus, antenna losses such as feed mismatch, feed loss, and waveguide and/or cable losses must also be considered. However, the radar engineer typically tabulates these losses separately for use in other radar calculations. Antenna directivity is an imposing calculation requiring volume integrals, but typical reflector pattern computation codes express the pattern in terms of directivity, and the antenna losses are generally accounted for separately. In the text that follows, reflector directivity is described based upon available "aperture gain" and associated radiation losses. This approach should be intuitive to the radar engineer.

Aperture-Field Method of Analysis. The aperture-field reflector analysis method is based upon ray-tracing principles and works well for symmetric, center-fed parabolic reflectors. With this method, the aperture field distribution is calculated in an x-y plane at $z > 0$ (Figure 12.1) by assuming coherent reflection of spherical radiation from a feed at the focal point. The field distribution is then used to compute the far-field radiation pattern. For more complex geometries, e.g., offset-fed reflectors, this method does not work as well as the more rigorous physical optics (PO) method that is described in Section 12.4.

For the simple center-fed focal feed example shown in Figure 12.1, the simpler aperture-field method of analysis is accurate and its application is simple and straightforward and makes the discussion of the radiation losses easier to follow. Using this method of analysis, the field amplitude on an x-y grid, F_{grid}, in the aperture plane is easily determined from the feed and space tapering:

$$F_{grid}(x,y) = F_{feed}(x,y,z)\ \frac{f}{r} = \frac{F_{feed}(x,y,z)}{1+\dfrac{x^2+y^2}{4f^2}} \tag{12.8}$$

where F_{feed} is the feed radiation pattern. For this case, with the feed at the focal point, the total distance traveled from the feed to the reflector and back to the aperture-plane is equal. This expression is deceptively simple and accurately accounts for the feed-reflector transformations and area projections.[1,2]

The resultant aperture field, $F(\hat{v})$, is then transformed to the far field using the spatial transformation:

$$F(\hat{v}) = \sum_{x,y} F_{grid}(x,y)\ \ e^{j\frac{2\pi}{\lambda}\left(x\,\hat{i}_x + y\,\hat{i}_y\right)\,\bullet\,\hat{v}} \tag{12.9}$$

where $\hat{v}$ is a unit vector in the direction of interest. Note that Eq. 12.9 is a simple 2D spatial summation analogous to that used for array antenna pattern calculations.

Taper Efficiency. In antenna design, aperture taper is used to lower sidelobes. The modest resultant loss and increased beamwidth is the price and one is usually willing to pay to obtain the desired sidelobe level. The loss associated with aperture taper is accounted for in the taper efficiency. However, the taper efficiency is not an ohmic loss where energy is dissipated, but is a redistribution of energy. In the case of a reflector fed by a horn, the taper distribution is determined by the feed horn pattern, the distance to the reflector, and the projected area in the direction of peak radiation. For a radially symmetric feed and reflector, where the field is radially symmetric and $F_{\text{grid}}(x,y) = g(r)$, the efficiency, η, is computed as

$$\eta = \frac{\left|\int g(r)\ 2\pi r\, dr\right|^2}{\int g^2(r)\ 2\pi r\, dr \int 2\pi r\, dr} \tag{12.10}$$

Spillover Loss. Spillover loss refers to feed power that misses, or spills over, the edges of the reflector. In the radar reflector design process, one typically adjusts the edge illumination to achieve a desired taper and sidelobe level resulting in modest spillover loss. Spillover loss is the feed power that is lost via radiation beyond the edges of the reflector. This loss can be computed as

$$\text{Spillover Loss} = \text{Total Feed Power}\left(1 - \frac{\text{Feed Power Incident on Reflector}}{\text{Total Feed Power}}\right) \tag{12.11}$$

For a radially symmetric feed and reflector, this calculation is straightforward.

Edge diffraction and resultant antenna backlobes are a related consequence of feed spillover and edge taper. For any given reflector design, edge illumination of the reflector will produce radiation behind the reflector due to diffraction. This diffraction may be thought of as re-radiation from the edge of the reflector that causes radiation lobes behind the reflector. For a center-fed reflector such as that shown in Figure 12.2, a primary backlobe will arise directly behind the reflector due to coherent addition of the edge diffraction currents. One common means of specifying this backlobe level is via the front-to-back F/B ratio, i.e., the ratio of the main beam and backlobe gains. The analysis of these reflector edge diffraction effects and associated F/B ratios for some common reflector geometries are described by Knop.[3]

Feed Blockage.[4,5] Many reflector systems suffer feed and/or feed-support blockage to some degree. For center-fed geometries, there will definitely be blockage because the feed is within the FOV of the reflector. A consequence of blockage is higher sidelobes, the levels of which depend upon the blockage area. Another consequence is loss due to blockage and depends upon the blocked electric field to main electric field ratio, E_b/E_m, which, in turn, is determined from the ratio of blocked power to total power on the main reflector, P_b/P_m, and ratio of the gain of the blocking object to the gain of the main reflector, G_b/G_m:

$$\left(\frac{E_b}{E_m}\right)^2 = \frac{P_b}{P_m}\ \frac{G_b}{G_m} \tag{12.12}$$

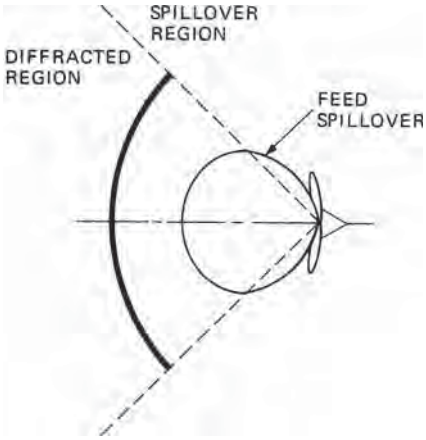


FIGURE 12.2 Graphical representation of spillover loss

Reflector taper is often approximated by a radial amplitude distribution[4]

$$g(r) = 1 - r^2 \tag{12.13}$$

where r is the radial distance normalized to the reflector radius and $g(r)$ drops to zero at the edge. The use of this taper function leads to a simple blockage loss equation:

$$\frac{P_b}{P_m} = \frac{g^2(0)\int_0^1 2\pi r\,dr}{\int_0^1 g^2(r)\;2\pi r\,dr}\;\frac{D_b^2}{D_m^2}$$
$$\frac{G_b}{G_m} = \frac{D_b^2}{\eta\,D_m^2} \tag{12.14}$$

where η is the efficiency, as given in Eq. 12.10. Using Eq. 12.14 and Eq. 12.15 leads to the simple expression for blockage loss typically found in the literature[5]:

$$\frac{E_b}{E_m} = 2\;\frac{D_b^2}{D_m^2}$$
$$\text{Blockage Loss} = \left(1 - \frac{E_b}{E_m}\right)^2 = \left(1 - 2\;\frac{D_b^2}{D_m^2}\right)^2 \tag{12.15}$$

where D_b and D_m are the diameter of the blockage and reflector, respectively.

The effective feed horn dimensions causing blockage may be different from the physical dimensions. If the walls are not tapered, such as shown in Figure 12.3, the effective size of the blockage hole can be larger than the projected obstacle area.

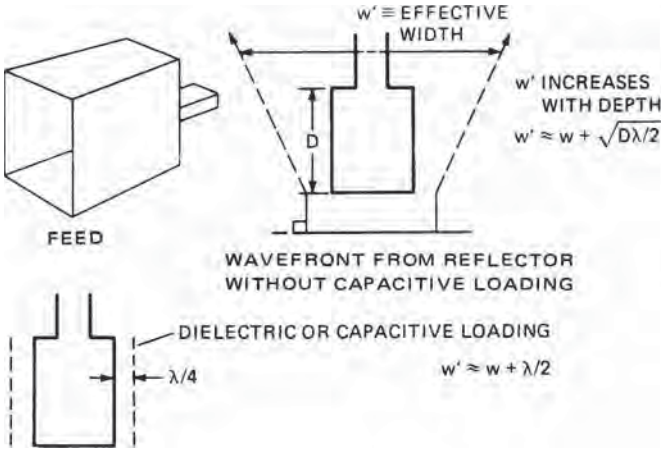


FIGURE 12.3 Capacitive loading to reduce feed horn blockage

For feeds with conducting walls, the effective H-plane width, w', is at least $\lambda/4$ and increases with depth D:

$$w' = w + \max\left\{\lambda/4, \sqrt{\lambda D/2}\right\} \tag{12.16}$$

where max indicates taking the larger of the two values. The increase in effective width can sometimes be reduced by adding capacitive loading, which can constrain the effective width to about a quarter wavelength wider than the physical width. The effective E-plane dimension is the same as the actual dimension.

Gain Optimization. Three of the principal reflector antenna losses, taper efficiency (loss), spillover, and blockage, have been explained in detail above. These design parameters (loss factors) affect one another and are often traded to optimize antenna performance, e.g., gain, sidelobes, etc. Other typical reflector losses, described later in this section, are resistive loss factors that simply reduce the antenna gain.

The design trades associated with taper efficiency, spillover, and blockage are illustrated by use of an example. Computed losses are shown in Figure 12.4 for a 20-wavelength diameter center-fed circular reflector with a 10-wavelength focal length, and a gaussian feed horn (a horn with a radiation pattern described by a gaussian function). The gaussian feed used in this sample analysis is a hypothetical feed with a radially symmetric pattern and very low spillover, and most typical feed horn patterns are well approximated by a gaussian feed model. The feed size determines the edge taper, i.e., the larger the feed, the greater the edge taper. The plot in Figure 12.4 shows how the three loss terms, and the total loss, vary as a function of feed pattern edge taper, that is, the feed power directed at the edge of the reflector, relative to the feed pattern peak. This is used here because the feed pattern is measured independent of the reflector. When the edge taper is low, virtually all the power strikes the reflector and the loss is insignificant. As taper decreases, there is more spillover, and feed power misses the reflector, increasing the loss. On the other hand, with too much taper, the taper efficiency is poor because the reflector is under illuminated.

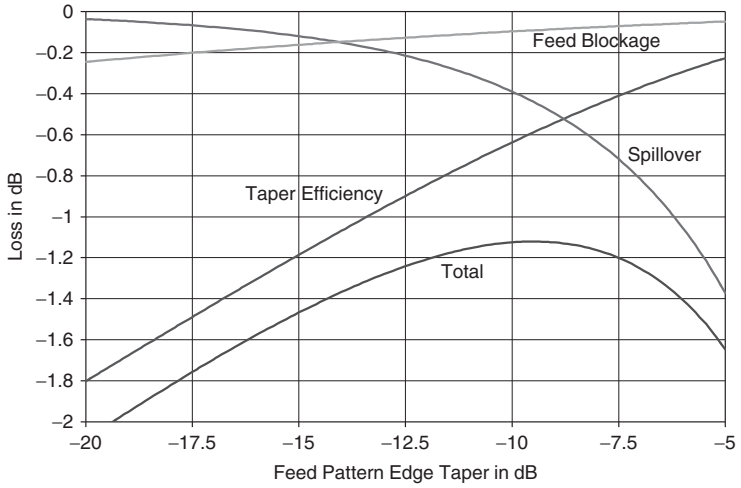


FIGURE 12.4 Taper efficiency, spillover, blockage, and total loss vs. feed pattern edge taper

Figure 12.4 demonstrates the gain optimization process with a hypothetical feed pattern, blockage, etc., and the total loss is 1.11 dB (77% efficiency) when a –9.5 dB edge taper is implemented. Although the plot of Figure 12.4 includes taper, blockage, and spillover losses, there are additional losses that must be included when assessing the overall aperture efficiency. These losses, e.g., feed blockage, surface reflection, feed mismatch, and resistive losses, etc., vary from system to system, but 0.8 dB is typical. With these additional losses, the overall loss becomes 1.91 dB or 64% aperture efficiency (typical for a single reflector system).

Sidelobe requirements must also be considered. As shown in Figure 12.5, for the same center-fed reflector system, the sidelobe level can be reduced by increasing the feed pattern edge taper.

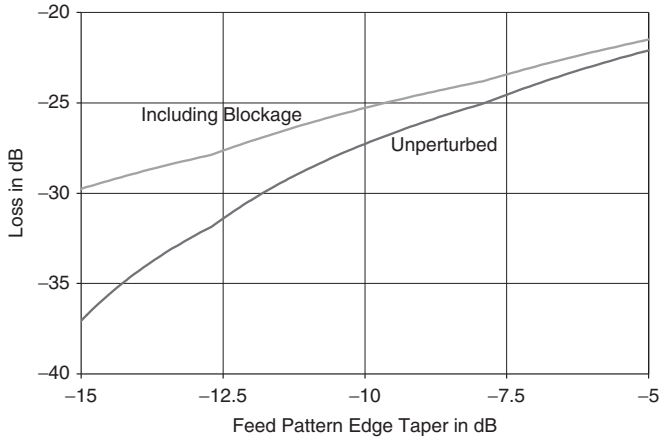


FIGURE 12.5 Sidelobe levels, with and without blockage, vs. feed pattern edge taper

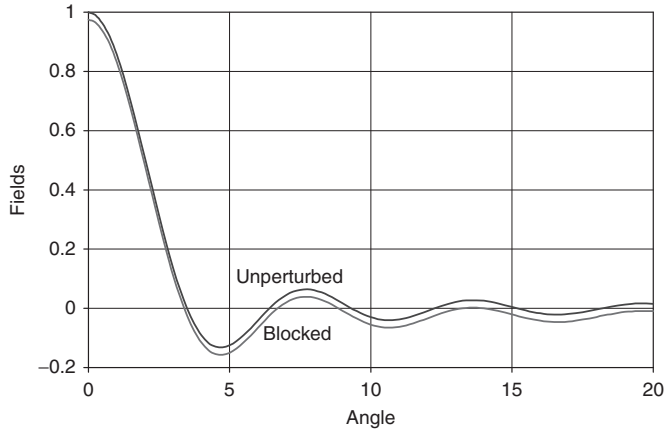


FIGURE 12.6 Unperturbed field and field including feed blockage

Figure 12.5 also shows that feed blockage increases the sidelobe level; for example, with a –10 dB edge taper, one obtains –25.5 dB sidelobes rather than –27.5 dB sidelobes. Figures 12.6 and 12.7 further illustrate this impact. This blockage effect can be modeled as a "hole" in the aperture that can be represented by a broad pattern with less gain. This blockage pattern is subtracted from unperturbed (unblocked) aperture fields, as shown in Figure 12.6.

The associated patterns, with and without blockage, are shown in Figure 12.7. The alternating large, small, large, and so on, sidelobe progression is characteristic of a blocked aperture.

Surface Leakage Loss. Many reflector surfaces are designed with a grid, a wire-mesh, or metalized fabric surface in order to minimize wind resistance, reduce weight, and/or enable stowage/deployment. Some common reflector mesh surface patterns are

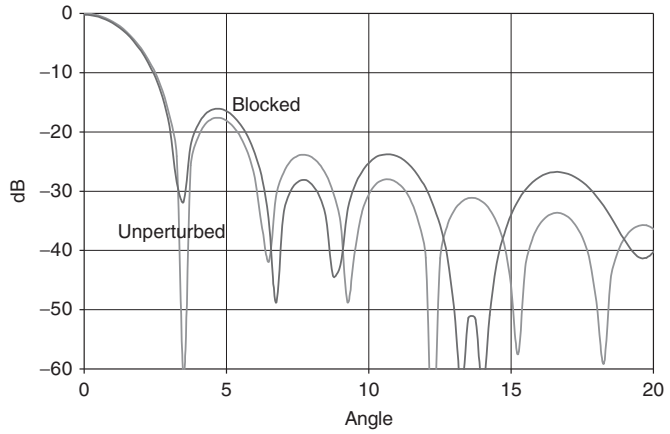


FIGURE 12.7 Unperturbed pattern and pattern including blockage

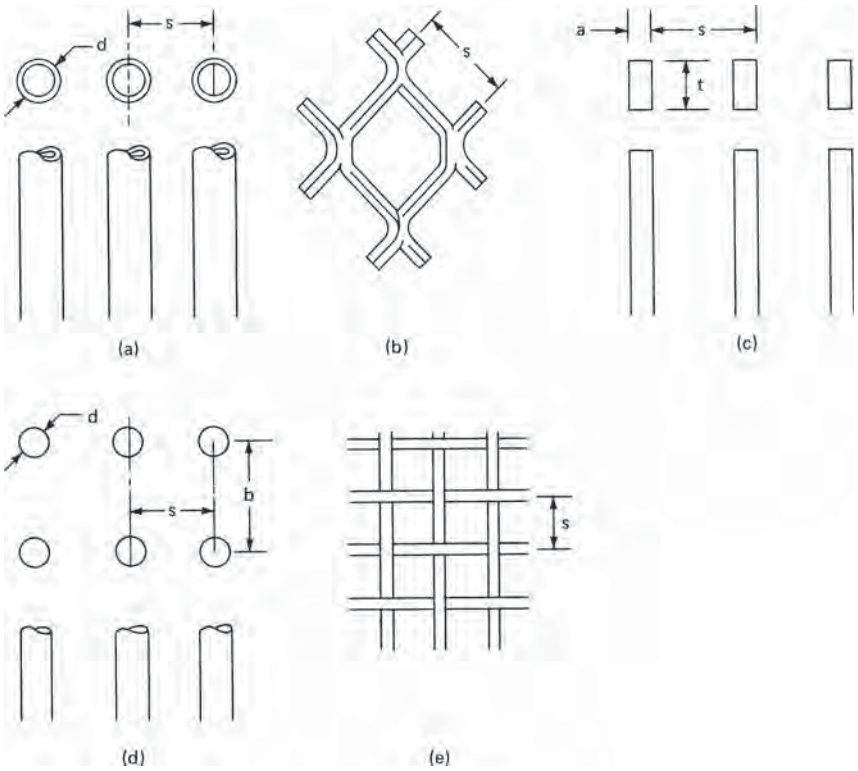


FIGURE 12.8 Common reflector surface materials for reduced wind resistance: (*a*) tubing, (*b*) expanded metal, (*c*) rectangular, (*d*) double layer, and (*e*) screen

shown in Figure 12.8. The mesh opening size is generally chosen to be as large as possible, but the gap, s, between conductors must be substantially less than $\lambda/2$ to cutoff and prevent significant transmission of electromagnetic energy through the surface. The attenuation, in decibels, through the passage depth t is approximately $27t/s$, plus fringing losses, which are approximately $27(\lambda/2s - 1)$.

The power passing through the reflector, or transmission loss, can be determined using a handy nomograph.[6] One selects the grid spacing and thread radius to achieve the transmission loss required. The resulting loss in antenna gain is termed *leakage loss*, and the relationship between leakage loss and transmission loss is plotted in Figure 12.9 for a conductive reflector (no ohmic losses). The leakage loss vs. spacing relationship for Figure 12.8*a* can also be computed using an equivalent circuit, with shunt susceptance, b, developed by Mumford,[7] the source of the nomograph[6]:

$$b = -\frac{\lambda}{s \ln\left(\dfrac{0.83}{1 - e^{-\pi d/s}}\right)}$$

$$\text{Leakage Loss} = \frac{1}{1 + b^2/4} \tag{12.17}$$

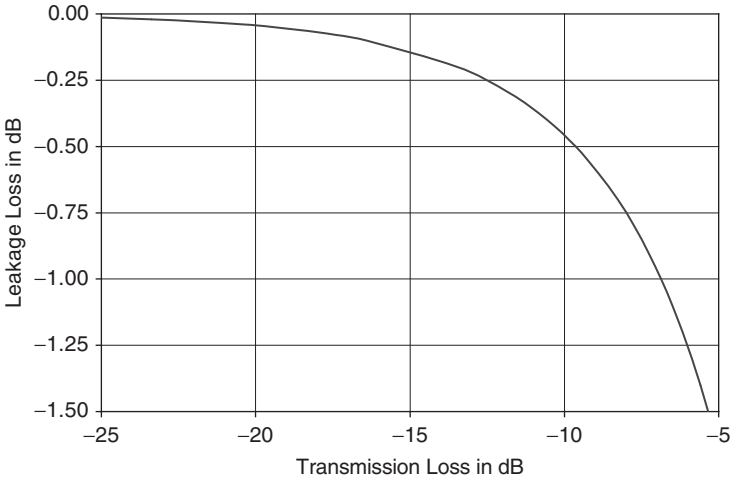


FIGURE 12.9 Reflector leakage loss vs. transmission loss

This nomograph[6] is frequently used for the simple configurations in Figure 12.8*a*. For other configurations and thread forms, the same equation can be used by computing an equivalent radius by matching the cross sectional areas.

One significant source of radiation in the back hemisphere is leakage through the reflector. This radiation level is determined from the difference between primary (feed horn) and secondary (reflector) gains and the leakage (loss) radiation through the mesh.

There are two considerations regarding the reflector surface leakage. The leakage loss affects the reflector gain directly. The radiation level in the back hemisphere is a function of the feed gain, reduced by the transmission loss through the reflector. Thus, the reflector designer must consider both the leakage loss and the transmission loss (Figure 12.9).

There are, however, additional backlobes that are distinct from the back radiation due to surface leakage. These additional lobes are due to reflector edge diffraction that adds coherently and typically causes a relatively strong main backlobe behind the reflector. Center-fed geometries, i.e., those with reflector edges equidistant from the feed, enable coherent addition of the diffracted energy directly to the rear of the reflector. For these geometries, the backlobe level is directly related to the edge illumination level.[3]

Surface Roughness Loss. All radar reflector antennas, especially those that are mechanically scanned or deployed, require careful consideration of mechanical design details. First, the reflecting surface must be designed and built so that it remains within close tolerances (typically $\pm 0.03\lambda$) even under dynamic operating and environmental conditions. Also, the feed must be accurately aligned with respect to the reflector. The feed-support structure and the reflector stiffening structure must maintain the feed location and surface dimensions while the antenna is being rotated. This dimensional stability must be maintained through wind loading, temperature variations, or other environmental factors to ensure that pattern performance is maintained.

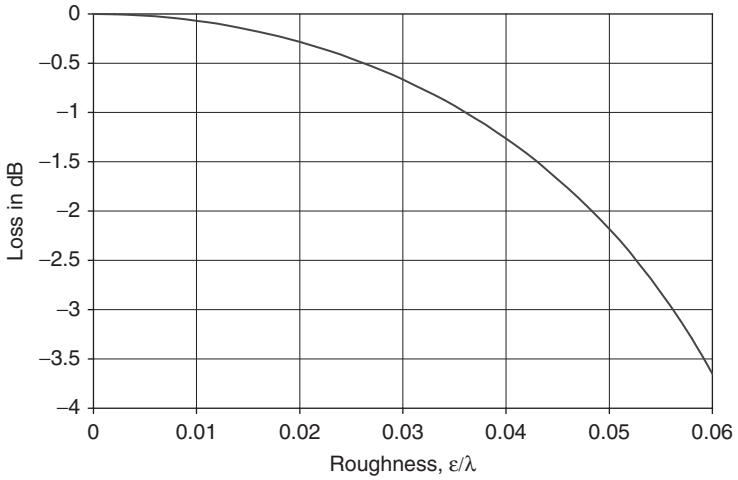


FIGURE 12.10 Surface roughness, ε/λ, vs. loss

Reflector antenna gain loss due to surface roughness has been discussed by John Ruze[8,9] using a statistical point of view. The loss of gain due to small roughness errors is approximately[9]

$$\frac{G}{G_0} \approx 1 - \overline{\delta^2} = 1 - \left(\frac{4\pi\varepsilon}{\lambda}\right)^2 \tag{12.18}$$

where $\overline{\delta^2}$ is the mean square phase error, ε is the effective rms surface error, λ is the wavelength, G_0 is the gain without phase error, and G is the gain with phase error.

The roughness loss effects embodied in Eq. 12.18 are captured in Figure 12.10 where gain loss vs. surface roughness (in lieu of phase error) is plotted. The plot shows that for 0.10 dB gain loss, the rms surface error must be less than 0.01λ. To maintain modest losses, surface errors must be tightly controlled.

Another consideration in mesh reflectors is systematic surface deformation. For many reflectors, a mesh is attached to a metal or composite backing structure. For spaceborne deployable reflectors, the mesh is typically a very sheer fabric-like membrane that is stretched across the surface and attached at a finite number of points to form the reflecting surface. In either case, there are surface distortions between the precisely controlled mounting points. This sets up a systematic error in the form of periodic cusps (see Figure 12.11). These errors typically cause pattern grating lobes because these error cusps are normally several wavelengths apart. The grating lobes are easily recognizable based on their

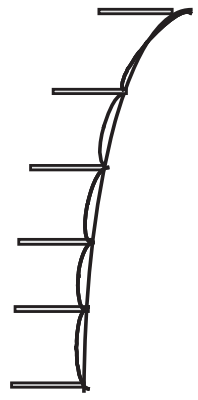


FIGURE 12.11 Systematic displacement between supports with a cusp between supports

well-defined angular spacing. For a grid with supports, and associated errors, a distance s apart, the grating lobe appears at

$$\theta = \arcsin(s/\lambda) \tag{12.19}$$

The grating lobe amplitude depends on the depth of the distortion (Figure 12.11), and is typically

$$\text{Grating Lobe} = \left(\frac{4\pi\varepsilon}{\lambda}\right)^2 \tag{12.20}$$

where ε is the depth of the cusp.

Feed Displacement. The permissible total error in the overall feed-reflector system alignment[10–12] is generally restricted to $\lambda/8$ or $\pm\lambda/16$. On the basis of this criterion, the resultant maximum deviation from the parabolic condition (feed at focus) would be $\pm\lambda/32$. Displacement of the feed along the focal axis (the z-axis in Figure 12.1) produces an even-order phase error on the aperture illumination. This results in modest beam broadening and some filling of the first null, but is normally not too detrimental.

Displacement of the feed normal to the z-axis produces an odd-order phase error in the aperture and causes higher sidelobes on one side of the beam. Furthermore, such displacement will cause the beam to mispoint. If the displacement is small, fixed, and known, one can typically calibrate out the fixed mispointing bias. However, if the displacement is random, say caused by vibration, the beam pointing error can be a problem. The reflector beam will be steered by θ radians if the feed is displaced off axis by an amount ε

$$\theta = \arctan(\varepsilon/f), \text{ radians} \tag{12.21}$$

where f is the focal length. So if the lateral feed displacement error is ε, the beam pointing error, $\Delta\theta$, is

$$\Delta\theta = \frac{\varepsilon/f}{1+(\varepsilon/f)^2}, \text{ radians} \tag{12.22}$$

Strut Blockage. Struts are used to support the feed, and for center-fed reflectors, typically form a tripod, as shown in Figure 12.12. Strut scattering is a complex phenomena that depends on strut size, strut geometry, field polarization, and other factors. However, in general, strut interference scattering will lie within a conical region about the strut axis. The strut is illuminated by a planar wavefront and one edge of the scattering cone lies on the reflector axis (see Figure 12.13). If the strut tilts at 30° to the axis, the scattering cone maximum angle is twice the tilt angle, or 60°. Thus, the scattering from three struts forms intersecting rings, as shown in Figure 12.14. This causes the most blockage at the pattern peak where the three rings intersect.

Polarization plays a significant role in strut scattering. For the example shown in Figure 12.12, strut #1 is parallel to the E-field and its blockage area is larger than its actual physical cross section. However, the lower struts, #2 and #3, are tilted 60° with respect to the E-field, and therefore appear smaller than their physical cross section.

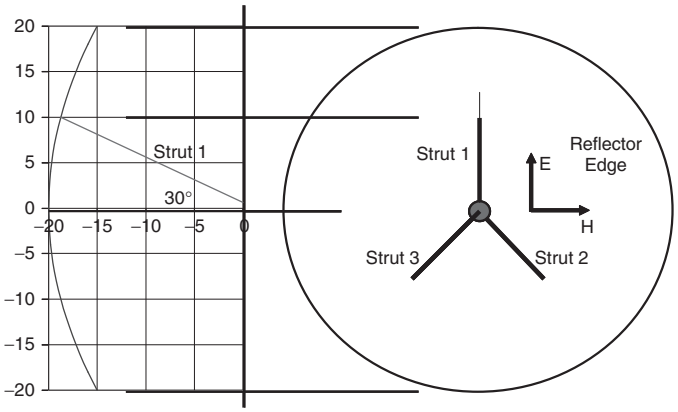


FIGURE 12.12 Center-fed reflector with tripod struts, side view, and axial view

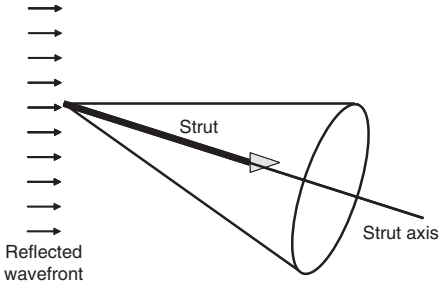


FIGURE 12.13 Strut 1 and maximum scattering cone

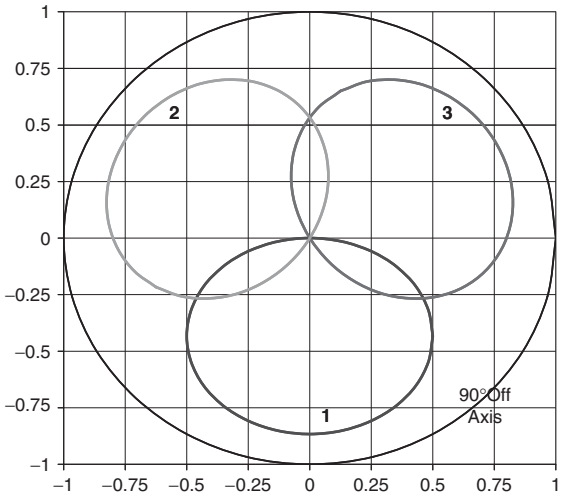


FIGURE 12.14 Axial view of 30° tripod strut blockage interference patterns

12.3 REFLECTOR ANTENNA ARCHITECTURES

Reflector antennas are built in a wide variety of shapes and sizes with a corresponding variety of feed systems to illuminate the surface(s), each suited to its particular application. Figure 12.15 illustrates the most common of these reflectors, each of which is described in some detail in the following subsections. The paraboloid in Figure 12.15*a* collimates radiation from a feed at the focus into a pencil beam, providing high gain and minimum beamwidth. The parabolic cylinder in Figure 12.15*b* performs this collimation in one plane but allows the use of a linear array in the other plane, thereby allowing flexibility in either steering or shaping of the beam in that plane. If beam shaping (not scanning) is the goal, an alternative method using a single feed, in lieu of a linear array, is shown in Figure 12.15*c*. Shaping of the surface along the vertical axis is used to spoil the beam shape in this plane, but because only the phase of the wave across the aperture is changed, there is less control over the beam shape than in the parabolic and/or phase cylinder shown in Figure 12.15*b* wherein the linear array may be adjusted in amplitude.

Very often the radar designer needs multiple beams to provide increased coverage or to determine angle. Figure 12.15*d* shows how multiple discrete feed locations produce a set of secondary beams at distinct angles. The additional feeds must be offset from the focus, and the resultant secondary beams will suffer some gain loss and distortion commensurate with the associated feed displacement. If an ESA feed is used, the reflector system can be designed to enable electronic beam scanning, albeit over a limited FOV. The use of modern ESA and ESA-like reflector feeds to achieve electronic scanning is discussed further in Section 12.4. An especially common multiple beam design is the monopulse antenna in Figure 12.15*e*, used for angle determination on a single pulse, as the name implies. In this instance, the second beam is normally a difference beam with its null at the peak of the first beam.

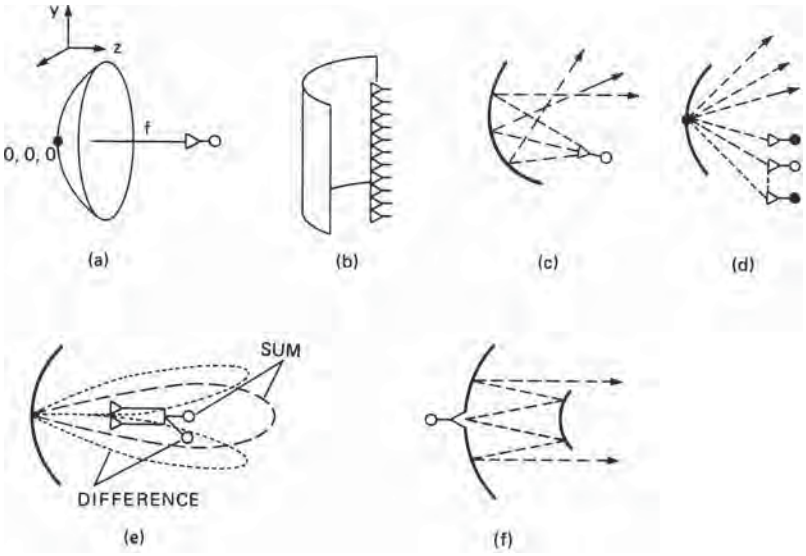


FIGURE 12.15 Common reflector antenna types: (*a*) paraboloid, (*b*) parabolic cylinder, (*c*) shaped, (*d*) stacked beam, (*e*) monopulse, and (*f*) Cassegrain

Multiple reflector systems, typified by the Cassegrain antenna shown in Figure 12.15*f*, offer one more degree of flexibility by shaping the primary beam and/or allowing the feed system to be conveniently located behind the main reflector. The symmetrical arrangement shown has significant blockage, but offset geometries mitigate feed blockage.

In modern reflector antenna design, combinations and variations of these basic types are widespread; however, design goals generally stress beam gain(s), shape(s), location(s), etc., while stipulating minimal losses and low sidelobe levels.

Paraboloidal Reflector Antennas. The theory and design of paraboloidal reflector antennas are extensively discussed in the literature.[13–17] A basic geometry is shown in Figure 12.16*a*, which assumes a parabolic reflector surface of focal length f with a feed at the focal point. It can be shown from geometrical optics considerations that a spherical wave emerging from the focal point, F, and incident on the reflector is transformed, after reflection, into a plane wave traveling in the positive z direction (Figure 12.16*b*).

Although reflectors are commonly illustrated with a round outline or rim and a central feed point, a variety of reflector aperture shapes are used in practice, as shown in Figure 12.17. Often, the azimuth and elevation beamwidth requirements differ, requiring an oblong aperture, as shown in Figure 12.17*b*-*e*.

If low sidelobe levels are required, feed blockage may become intolerable, requiring the use of an offset feed (Figure 12.17*c*). The feed is still generally located at the focal point, but the reflector is realized via use of a different portion of the parabola. For offset-fed reflector configurations, the focal axis generally does not intersect the reflector surface. Feeds for offset-fed reflectors are generally aimed close to, but slightly beyond, the center of the reflector area to account for the larger space taper (spreading loss) on the far side of the reflector. This generally results in a slightly unsymmetrical aperture illumination.

The corners of most paraboloidal reflectors are frequently rounded (not shown) or mitered (Figure 12.17*d*) to reduce extraneous surface area and/or minimize the torque required to turn the antenna. Because these corner regions are generally weakly

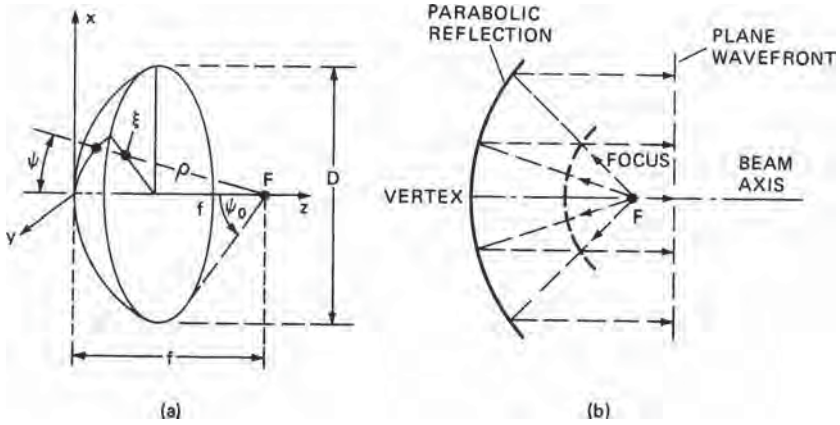


FIGURE 12.16 Geometrical representation of a paraboloidal reflector: (*a*) geometry and (*b*) operation

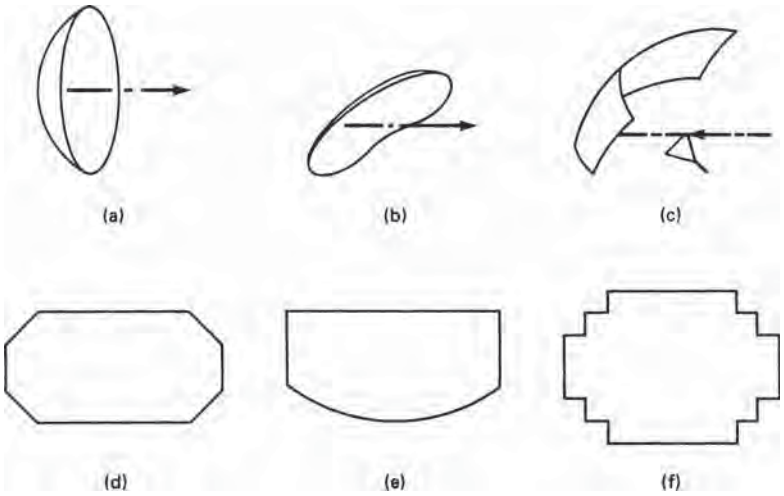


FIGURE 12.17 Paraboloidal reflector aperture shapes or rims: (*a*) round outline/rim, (*b*) oblong, (*c*) offset feed, (*d*) mitered corner, (*e*) square corner, and (*f*) stepped corner

illuminated by the feed, their removal generally has little impact on the gain. However, circular and elliptical outlines/rims will produce modest sidelobes in both principal and nonprincipal planes. If very low sidelobes are specified in nonprincipal planes, it may be necessary to maintain square corners, as shown in the upper part of Figure 12.17*e*.

Parabolic-Cylinder Antenna.[18,19] It is quite common that either the beam must be steerable or shaped in only one plane, either azimuth or elevation. A parabolic cylindrical reflector fed by a linear array feed can accomplish this at moderately higher cost.

The parabolic cylinder antenna can be applied to achieve a precisely shaped beam from a common aperture. The AN/TPS-63 (Figure 12.18) uses a vertical array to provide fine control of the elevation pattern with a single elevation column feed array and so is very cost effective. The elevation beam-shaping incorporates a steep beam slope at the horizon to allow radar operation at low elevation angles without degradation from ground reflections. The TPS-63 produces a much sharper slope at the horizon than a shaped reflector of equal height. The array feed enables superposition of beams close to the aperture normal, thereby enabling very high taper efficiency (near full aperture gain).

FIGURE 12.18 AN/TPS-63 parabolic-cylinder antenna (*Courtesy Northrop Grumman Corporation*)

The basic parabolic cylinder geometry is shown in Figure 12.19. The reflector surface is defined as

$$z = \frac{y^2}{4f} - f \tag{12.23}$$

where z is the distance from the focal plane, f is the focal length of the cylindrical reflector, and y is the horizontal dimension. For the parabolic cylinder, the surface does not vary with x, the height. The feed is generally on the focal line, and in many ways, the design of parabolic cylinder reflectors is similar to that of paraboloidal reflectors. One significant difference is that the feed energy diverges cylindrically rather than spherically, and so the feed power density falls off as $1/\rho$ rather than $1/\rho^2$.

The height of the parabolic cylinder (Figure 12.19*a*) must allow for beamwidth, shaping, and steering of the linear feed array. When the line source steers at angle θ from broadside, the primary beam from the source lies on a conic, and the intercepts at the upper right and left corners of the reflector are farther up than in the center, as shown in Figure 12.19*b*. Therefore, the corners of parabolic cylinder reflectors are seldom rounded in practice.

Shaped Reflectors. Fan beams with a specified shape are required for a variety of reasons. A common elevation shaped beam requirement is to provide equal echo signal power on targets at constant altitude. If the transmit and receive beams are identical, and if secondary effects are ignored, this can be achieved with a power radiation pattern proportional to $\csc^2\theta$, where θ is the elevation angle.[13] In practice, the well-known cosecant-squared pattern is typically modified to account for the curvature of the earth and the characteristics of sensitivity time control (STC).

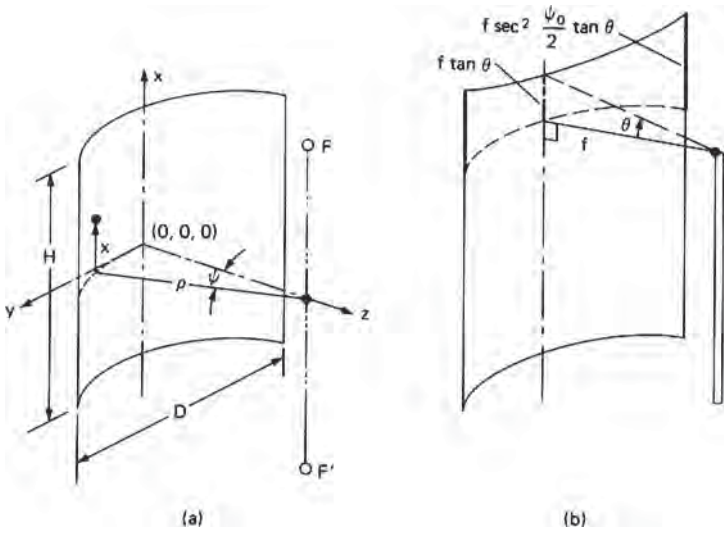


FIGURE 12.19 Parabolic cylinder: (*a*) geometry and (*b*) surface extension enables line source steering.

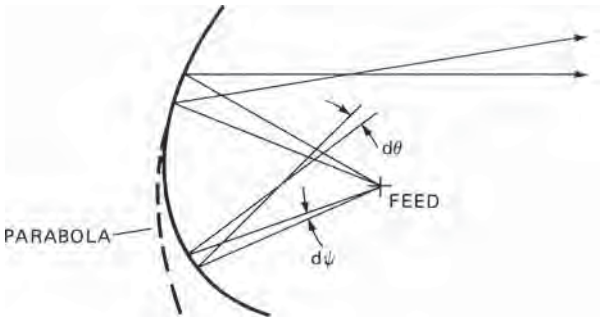


FIGURE 12.20 Reflector shaping

A relatively simple way to shape the beam is to shape the reflector, as Figure 12.20 illustrates. Each portion of the reflector is designed to reflect a portion of the incident energy in a different direction, and to the extent that geometric optics is valid, the power density at that angle is the integrated sum of the power density from the feed across that portion of the reflector. Silver[13] graphically describes a procedure for determining the reflector contour for a cosecant-squared beam. However, computer software packages now exist that enable synthesis of arbitrary beam shapes via application of iterative optimization techniques in conjunction with physical optics–based pattern computations. These analysis methods and software packages are summarized in Section 12.5.

Offset-fed parabolic reflectors are often used to mitigate feed blockage. However, reflector shaping can also be used to mitigate blockage and is sometimes applied to redirect the reflected energy away from the feed, as shown in Figure 12.20. Figure 12.21 illustrates how shaping can be applied to virtually eliminate blockage even though the feed appears to be within the FOV of the reflector.

The ASR-9 antenna (Figure 12.22), found at major airports, typifies shaped reflector antenna design. The elevation beam shape is tailored using a computer-aided design process, and azimuth sidelobes are lowered by offsetting the feed so there is no blockage. Two linear or circularly polarized elevation beams are produced by using two feedhorns. One beam is close to the horizon for close in targets, and the second is higher in elevation, so it receives less ground clutter for long range targets.

One typical characteristic of shaped reflectors is lower aperture efficiency. Generally, the surface shaping process spoils the aperture phase and broadens the beam thereby inherently reducing the aperture efficiency. However, this sacrifice in aperture efficiency is generally accepted by the radar designer when beam shaping is desired or required.

The second antenna atop the ASR-9 reflector (Figure 12.22) provides an independent tracking system. It is an Air Traffic Control Radar Beacon System array antenna that transmits and receives narrow azimuth sum, difference, and guard beams, all shaped in elevation. It requires a transponder aboard the aircrafts targeted, as it has low gain.

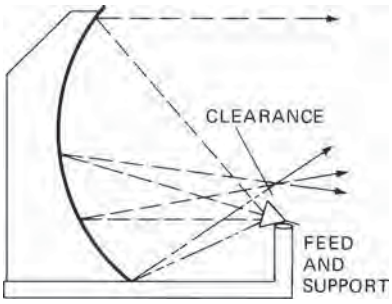


FIGURE 12.21 Elimination of blockage

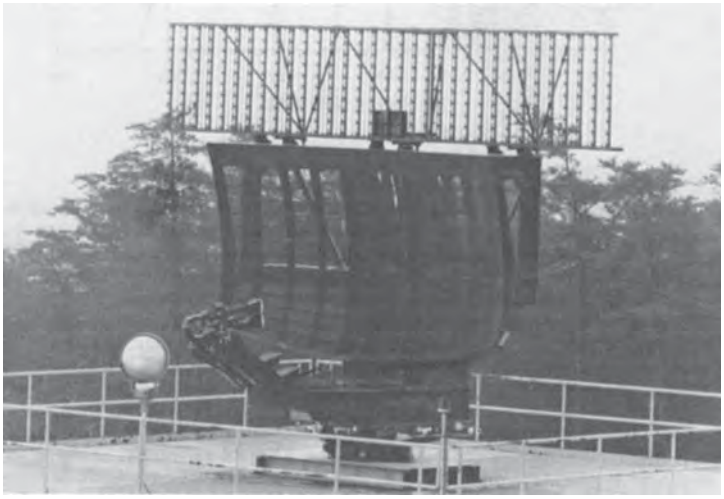


FIGURE 12.22 ASR-9 shaped reflector with offset feed. Air Traffic Control Radar Beacon System (ATCRBS) array is mounted on top. (*Courtesy Northrop Grumman Corporation*)

Multiple-Reflector Antennas.[20–30] There are both difficulties and advantages associated with adding a secondary, or subreflector, to a paraboloidal reflector system. The shape of the subreflector determines how the power will be distributed across the primary reflector and thereby provides some control over amplitude in addition to phase in the aperture, making it possible to produce very low spillover or to produce a specific low-sidelobe distribution. By suitable choice of shape, the apparent focal length can be enlarged so that the feed size is practical or even feasible. This is sometimes necessary for monopulse operation.

The Cassegrain dual reflector antenna (Figure 12.23), derived from optical telescope designs, is the most prevalent dual reflector configuration. Figure 12.23*a* shows a small subreflector between the feed and parabolic main reflector. The feed illuminates the hyperboloidal subreflector, which in turn illuminates the paraboloidal main reflector. The feed is placed at one focus of the hyperboloid, and the paraboloid focus is coincident with the second focus of the hyperboloid. The use of a subreflector also allows the feed to be located behind the main reflector and closer to the transmitter and receiver in order to minimize transmission line losses. Furthermore, if the feed is located behind the main reflector, the center of gravity will be biased closer to the main reflector vertex thereby simplifying the design of both the structure and the gimbal (precision mechanical positioning system).

The Gregorian dual reflector antenna system (not shown in Figure 12.23) is similar to the Cassegrain, but it uses an ellipsoidal subreflector in lieu of a hyperboloid resulting in a lengthening of the reflector system (along the focal axis).

The parameters of the Cassegrain reflector[20] are related by the following expressions:

$$\begin{aligned} \tan \psi_v / 2 &= 0.25 D_m / f_m \\ 1/\tan \psi_v + 1/\tan \psi_r &= 2 f_c / D_s \\ 1 - 1/e &= 2 L_v / f_c \end{aligned} \tag{12.24}$$

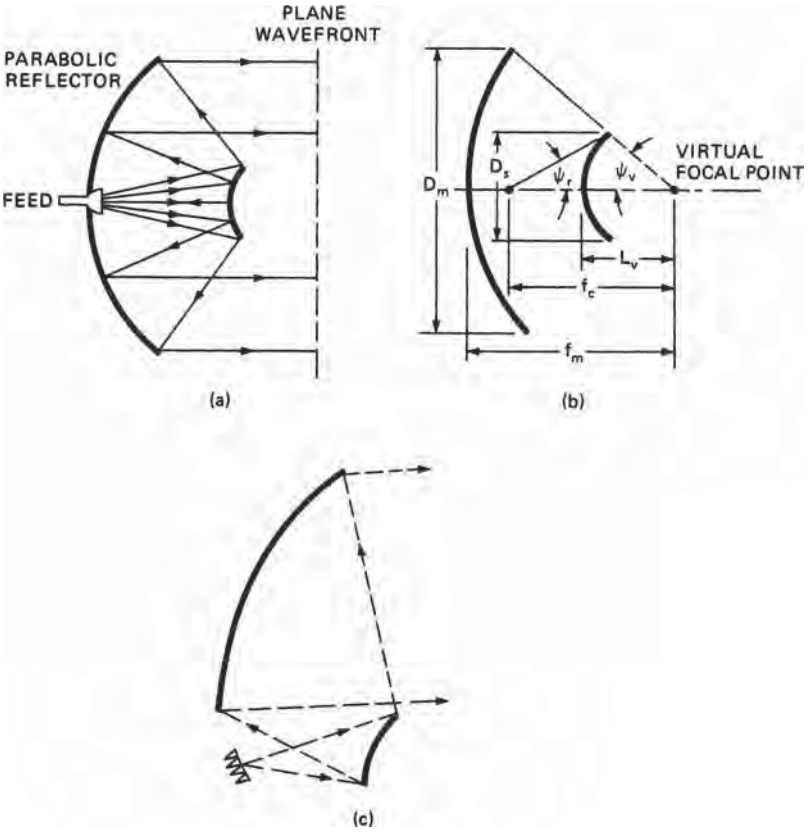


FIGURE 12.23 Cassegrain dual reflector antenna systems (the larger surface is the main reflector and the smaller surface is the subreflector): (*a*) ray optics (*b*) typical axial configuration,[20] and (*c*) offset configuration

where the eccentricity e of the hyperboloid is given by

$$e = \sin[(\psi_v + \psi_r)/2]/\sin[(\psi_v - \psi_r)/2]$$

The equivalent-paraboloid[20] concept is a convenient method of analyzing the radiation characteristics using a single reflector "equivalent" model. This method utilizes a paraboloid of equal diameter, but a larger focal length to model the dual-reflector Cassegrain system. The equation

$$f_c = D_m/(4\tan(\psi_r/2)) \tag{12.25}$$

defines the equivalent focal length, and the focal length ratio or magnification m is given by

$$m = f_c/f_m = (e+1)/(e-1) \tag{12.26}$$

The magnification m is a useful metric in that it provides a measure of the reduction in size/length along the focal axis that is enabled by use of the Cassegrain reflector system in lieu of a single parabolic reflector system. The feed is designed to produce suitable illumination within subtended angles $\pm\psi_r$ associated with the longer focal length f_c.

Aperture blocking can be large for center-fed Cassegrain antennas. The blockage can be minimized by choosing the diameter of the subreflector to be equal to that of the feed.[20] This occurs when

$$D_S = \sqrt{2f_m\ \lambda/k} \tag{12.27}$$

where k is the ratio of the feed-aperture diameter to its effective blocking diameter. Ordinarily k is slightly less than 1. For linearly polarized radar applications, aperture blocking can be significantly reduced by using a polarization-twist reflector and a subreflector made of parallel wires.[24] This twist reflector design enables a 90° rotation of the polarization such that the polarization of the beam upon reflection from the main reflector is orthogonal and transparent to the gridded subreflector.

Blockage can also be eliminated by offsetting both the feed and the subreflector (Figure 12.23*c*). With blockage and supporting struts and spillover virtually eliminated, this geometry is useful for very low sidelobe applications.[24]

As described earlier, the aperture efficiency of single reflector systems is maximized by balancing the feed taper and the feed spillover and minimizing other losses, but is typically 55–65%. However, dual reflector systems (e.g., Cassegrain) have an additional degree of freedom and surface shaping can be applied to decrease the taper loss and enable aperture efficiencies in excess of 70%.[25,26]

A different type of dual reflector system that is particularly well-suited to applications where limited electronic scanning is required is the so-called confocal reflector system[27–30] shown in Figure 12.24. This system employs two parabolic reflectors, a main and a sub, that share a common focal point. The optics of this system is designed such that a plane wave source, say from an array, is first converted to a spherical wave at the subreflector. Then, upon reflection from the subreflector, the feed energy converges at the common focus and diverges again as a spherical wave before finally reflecting from the main reflector.

The confocal system has several interesting properties, tied to the magnification factor M:

$$M = f_M / f_S \tag{12.28}$$

where f_M and f_S are the focal lengths of main and subreflectors, respectively. The first property shows that the system is essentially a feed source magnifier. The reflector gain, G_r, is described by

$$G_r \approx G_f \times M^2 \times \cos(\theta_r) \tag{12.29}$$

where G_f is the gain of the feed array and θ_r is the scan angle of the reflector (secondary) beam. The second property defines the scanning. The reflector scan angle, θ_r, is defined by the following equation:

$$\theta_r \approx \theta_f/M \tag{12.30}$$

where θ_f is the feed scan angle. For example, if the magnification factor is 10 and the ESA feed array is scanned 30°, the reflector beam will scan approximately 3°.

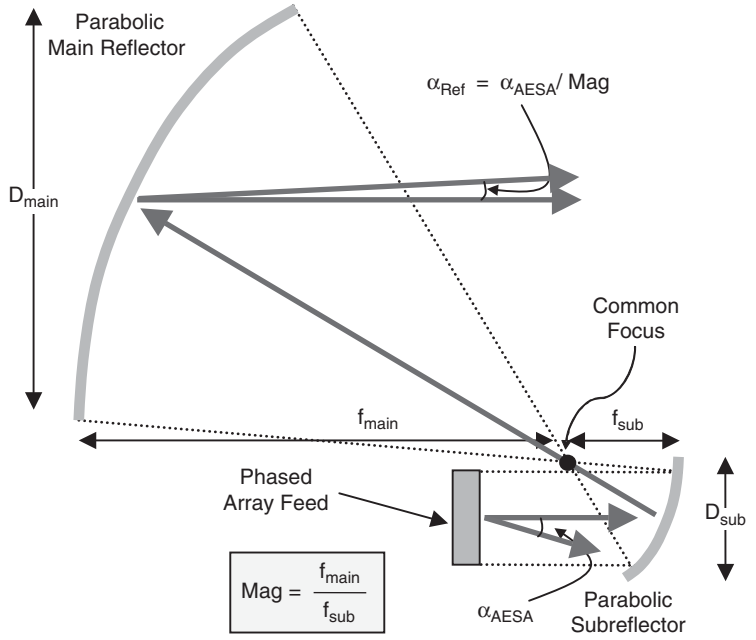


FIGURE 12.24 Confocal dual reflector antenna

Scan aberrations lead to moderately higher scan losses and slight beam scan angle deviations from those predicted by Eqs. 12.29 and 12.30 when ideal ESA plane wave excitations, i.e., linear phase slope, are applied. However, the amplitude and phase controls in the ESA allow for scan aberration compensation via conjugate matching of the aperture fields.[29] Thus, the use of an ESA feed enables recovery of scan losses resulting from aberrations in the system.

Confocal reflector architectures have been the subject of study and analysis for many years and numerous demonstration systems have also been developed.[25–28] Although confocal reflectors have not been adopted for significant usage in operational systems, these reflector systems are likely to be part of some next-generation radars due to the relatively recent maturation of enabling ESA technology.

Spherical Reflectors.[31–34] The spherical reflector[31] is sometimes used for applications requiring scanning or multiple beams over very wide angles. Its design is based upon the fact that, over limited angular regions, a spherical surface viewed from any point halfway between the center of a sphere and its surface is nearly parabolic. This means that if a feed is moved along an inner spherical surface of constant radius $R/2$, where R is the radius of the spherical reflector, the secondary beam can be steered. The range of beam steering is limited by the size of the spherical reflector, i.e., the portion of a full sphere realized by the reflector. The scanning capability can be implemented via use of either a single movable feed or an array with switchable feeds.

Self-blockage (reflector blocking itself) is another potential limiting factor in spherical reflector systems. However, 360° of azimuth steering can be accomplished via a polarization design scheme similar to the polarization twist subreflector described

above[24] for the Cassegrain system. In this design, the feed polarization is tilted 45° and the reflector is formed of conducting strips parallel to the polarization. However, the conducting strips on opposing sides of the reflector are twisted 90°, thereby enabling transmission of the reflected energy. This type of antenna is known as a *helisphere*.

If wide angle scanning in only one plane, i.e., azimuth or elevation, is required, a similar design called the *parabolic torus*[32] is more applicable. The parabolic torus reflector is a surface that is spherical in one plane (azimuth or elevation) and parabolic in the other plane. This design takes advantage of both the wide scanning enabled by the spherical shaping and the high aperture efficiency enabled by the parabolic shaping. The height (elevation) dimension of the reflector is set by the required elevation beamwidth.

The parabolic torus has seen application in various radars including the original BMEWS system[33] and the SPS-30 and SPQ-9B[34] systems.

12.4 REFLECTOR FEEDS

Whereas phased array antennas are frequently chosen for radar system designs, reflector antennas were once the dominant antenna design choice for medium- to high-gain radar apertures. Obviously, the cost of a single feed horn and metal reflector is much less than the same size array with many individual elements and associated phase shifters, amplifiers, receivers, etc. Consequently, many radars currently in the field use reflector antennas. Furthermore, reflector antenna designs are used in modern radar designs where moderate scan rates and scan volumes are required or low cost is essential.

Many legacy radars employ reflector antennas with single feed elements or a cluster of feeds (i.e., an array) with a fixed beam RF combiner or divider network. Figure 12.25 shows some typical single-feed reflector configurations utilizing horn, waveguide, and dipole feeds. Figure 12.26 shows a few common types of flared horn feeds.

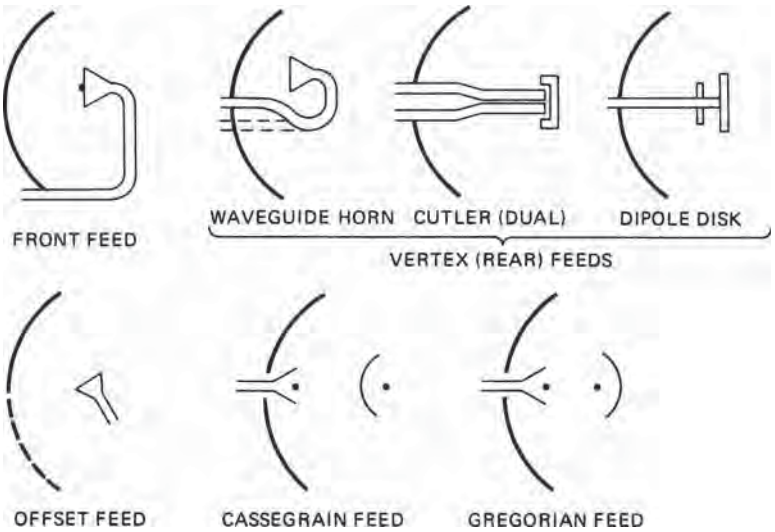


FIGURE 12.25 Some typical reflector antenna configurations

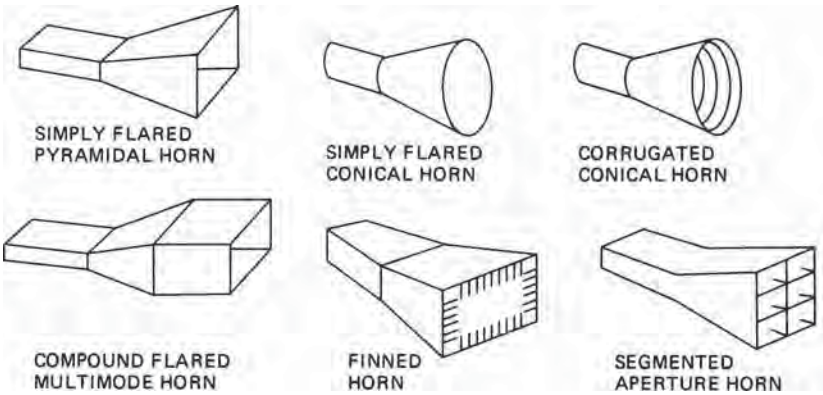


FIGURE 12.26 Various types of horn feeds for reflector antennas

Other types of feeds such as dipoles, microstrip patches, notches, etc., are sometimes used, but for single-feed reflector implementations, flared waveguide horns are the most common. The feed (in conjunction with the reflector) must also satisfy the antenna polarization requirements and handle required peak and average power levels under all operational environments. Other feed design considerations include the operating bandwidth and the potential implementation of any additional modes/patterns, e.g., difference or squinted beams.

For single-feed radar reflector antennas such as those depicted in Figures 12.25, mechanical scanning is generally achieved via a gimbal (a precision mechanical pointing system). Radar antenna gimbal designs vary greatly depending upon scan rate, FOV, tracking requirements, antenna size, mass, etc.

Basic Feeds.[35,36] For radars requiring a simple pencil beam, basic single mode waveguide horn feeds such as pyramidal (TE_{01} mode) and conical (TE_{11} mode) horns are widely used. Single-mode, flared horns will provide linearly polarized pencil beams and will generally handle high power. For more demanding applications requiring tracking modes, polarization diversity, high beam efficiency, or ultra-low sidelobes, feed designs become correspondingly more complex. For such applications, segmented, finned, multimode, and/or corrugated horns, as illustrated in Figure 12.26, are often used. Multimode feeds enable realization of difference mode patterns with a compact, single feed-horn and, as such, are especially useful for tracking applications.

Monopulse Feeds.[37–40] Monopulse is frequently used in radar tracking and surveillance systems to either keep the beam pointed on the target (tracking) or to accurately measure the angle to the target (surveillance).[37–40]

Amplitude comparison monopulse systems, illustrated in Figure 12.27, use the sum of the two feed

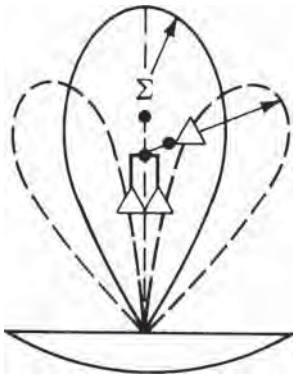


FIGURE 12.27 Amplitude comparison monopulse antenna showing sum and difference beams

outputs to form a high-gain low-sidelobe beam and the difference of the two squinted beams to form a precise, deep null at boresight. The sum beam is used on both transmit and receive to detect the target. The difference beam is receive only and provides angle determination. In most applications, both azimuth and elevation difference beam ports are implemented. Figure 12.27 illustrates the concept of amplitude comparison monopulse. A feedback loop minimizes the received difference beam signal by mechanically steering the antenna to keep the null (and the corresponding sum beam peak) on target.

There are numerous ways to realize amplitude monopulse beams in a reflector antenna design, but most designs generally fall into two classes; (1) multifeed and (2) multimode. Multifeed designs use combining networks to generate differential feed distributions. In its simplest form, an (az/el) multifeed monopulse feed array can be realized as a four-element feed, as shown in Figure 12.28. However, some applications use more feed elements to further tailor the distribution to improve efficiency and/or difference beam slope.

If the reflector is illuminated with a simple four-element feed, a conflict generally arises between the goals of high sum-beam efficiency and high difference-beam slope from the monopulse comparator. The former requires a small overall feed size, whereas the latter requires large individual feeds (Figure 12.29). Numerous design methods have been devised to overcome this problem, as well as the associated high difference pattern sidelobes. These methods either use different sets of feed elements for the sum and difference beams or apply different array amplitude/phase weightings for each of the beams. If horn feed elements are used, one approach is to oversize the feeds to enable multimode excitation for the sum beam as described by Hannan.[40] A comparison of some common monopulse feed configurations is included in Table 12.2.

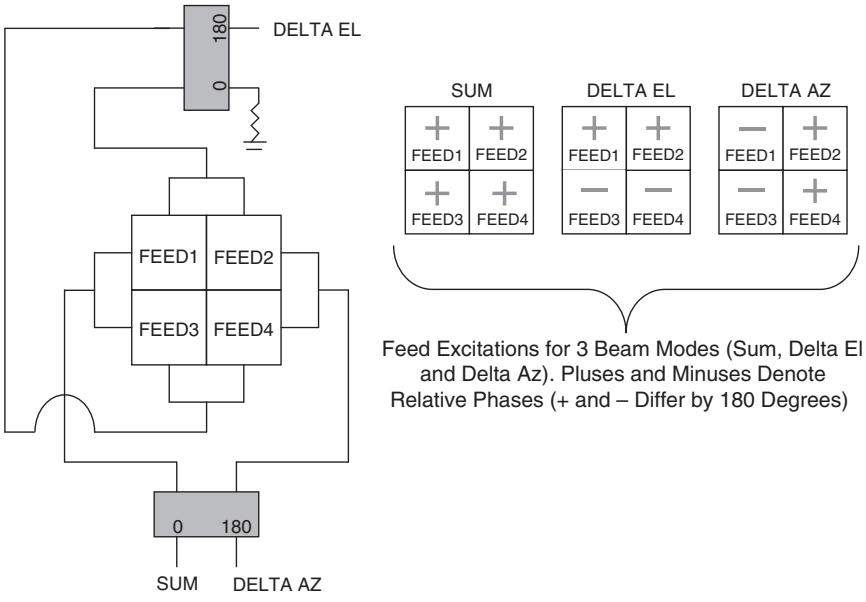


FIGURE 12.28 Four-element monopulse feed with dum, delta az, and delta el ports

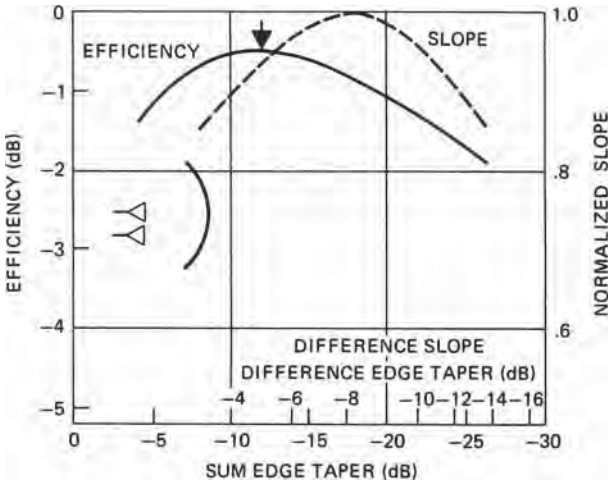


FIGURE 12.29 Plots of sum-beam efficiency and difference-beam slope as a function of associated edge tapers (*H*-plane shown)

Array Feeds. A single feed at the focal point of a parabola forms a beam parallel to the focal axis. Additional feeds displaced from the focal point form additional beams at angles from the axis (see Eq. 12.21). Thus, one can employ a multiple feed array with appropriate electronics to feed a reflector and provide either multiple displaced beams or electronic beam switching, i.e., beam scanning to discrete angles. A reflector system such as this can effectively provide electronic scanning over a narrow FOV. However, this type of array-fed reflector architecture does suffer from one drawback. A parabola converts a spherical wave into a pure plane wave only when the source (feed) is at the focus. If the source (feed) is displaced from the focus, the reflected wave is somewhat distorted and this results in gain loss and beam shape distortion. Figure 12.30 shows the effect of this distortion on the pattern of a typical center-fed reflector as the feed is moved off-axis.

TABLE 12.2 Monopulse Feedhorn Performance

	H plane		*E* plane	Sidelobes, dB		
Type of Horn	Efficiency	Slope	Slope	Sum	Difference	Feed shape
Simple four-horn	0.58	1.2	1.2	19	10	
Two-horn dual-mode	0.75	1.6	1.2	19	10	
Two-horn triple-mode	0.75	1.6	1.2	19	10	
Twelve-horn	0.56	1.7	1.6	19	19	
Four-horn triple-mode	0.75	1.6	1.6	19	19	

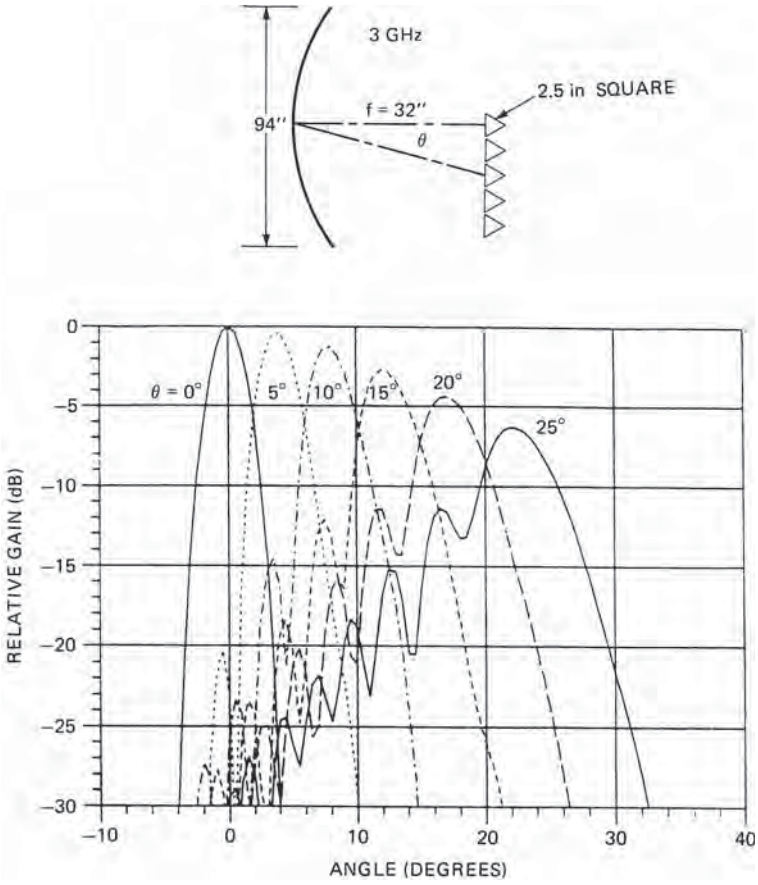


FIGURE 12.30 Patterns for off-axis feeds

A phased array (ESA) type feed covering the focal region (planar or curved in shape), such as that shown in Figure 12.31*b*, can provide improved capability. It offers two enhancements that extend the capability above and beyond that of a multiple feed array with switched beams. The phased array fed reflector can transmit or receive beams over continuous angles within the FOV whereas the multiple feed array is limited to discrete beam positions. It also provides greater aperture efficiency because it allows for adjustment of the element amplitudes and phases to reduce scan losses due to aberrations. If we view the reflector as a collector of parallel rays from a range of angles covering the FOV and examine the converging ray paths (Figure 12.31*a*), it is evident that an appropriately sized feed region can be found that intercepts most of the energy. If the phased-array feed array is appropriately designed, beam distortion and scan losses (such as those shown in Figure 12.30) can be effectively eliminated. This sort of design approach wherein the feed array amplitude/phase distribution is "matched" to the focal plane fields is often referred to as *conjugate field matching*.[41]

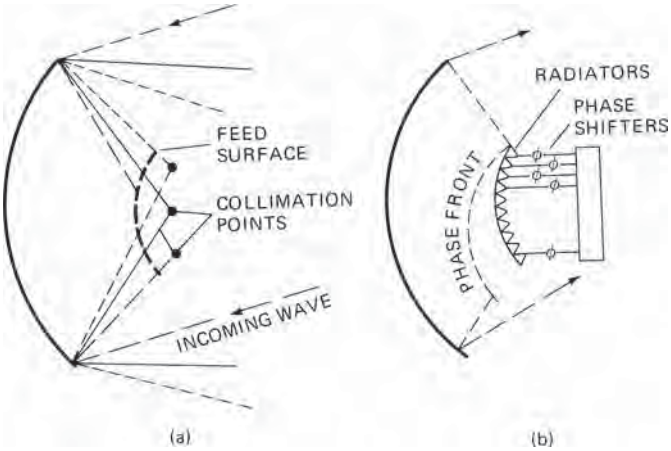


FIGURE 12.31 Extended feed region improves sidelobes of offset feed (used on ARSR-4): (*a*) ray geometry and (*b*) curved feed

Feed arrays, such as that illustrated in the ARSR-4 reflector antenna shown in Figures 12.31 and 12.32, have also been used to enable shaped beams and low sidelobe patterns. Passive combiner networks are used to generate multiple receive beams stacked in elevation and a single transmit beam. The receive beams require low azimuth sidelobes. A conventional beamforming approach would incorporate an array at the focal plane with a single beam per feed; however, the associated phase distortion due to displacement causes poor azimuth sidelobes. To correct this problem, the feed array is placed forward of the azimuth focal point, enabling compensation and sidelobe improvements via use of multiple feeds per beam with appropriate feed phasing (Figure 12.31*a*). Two different focal lengths are used in the reflector, one for elevation and a longer one for azimuth. The feed is on a curved surface, optimized via ray tracing, and is forward of the azimuth focal point. For each row, the amplitude and phase is optimized for a beam with low azimuth sidelobes. The elevation patterns for each single row have poor sidelobes, but several rows are used for each receive beam thereby improving the elevation sidelobes. For transmit, all 24 rows (the whole array) are used. On receive, 9 groups of rows are used to form 9 receive beams.

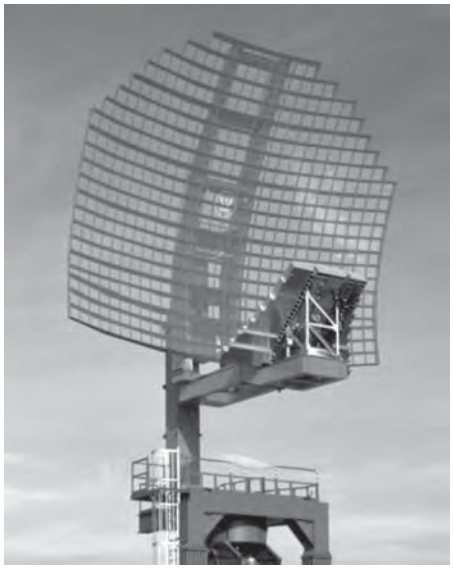


FIGURE 12.32 ARSR-4 low-sidelobe reflector with offset array feed (*Courtesy Northrop Grumman Corporation*)

12.5 REFLECTOR ANTENNA ANALYSIS

Reflector antenna analysis methods can generally be lumped into three classes or categories: (1) Physical Optics (PO), or induced current methods, (2) Geometrical Optics (GO) methods, with and without diffraction terms, and (3) rigorous, or full-wave, methods.

Physical Optics (PO) Reflector Analysis. The Physical Optics (PO) method is commonly employed for the most rigorous reflector analysis due to its accuracy. It incorporates the field properties of the feed, and models the resultant fields from the reflector, thus enabling the computation of the cross polarization properties. Furthermore, the method is more accurate for feeds that are not at the focal point and reflectors that are not parabolic. There are many good references that describe the theory behind the method and how it is applied to the analysis of reflector antennas.[42–44] PO is a very general "high frequency" analysis method that generally provides high fidelity pattern predictions for most reflector systems as long as the reflector dimensions are large, say, greater than about five wavelengths in both dimensions. An overview of PO is provided here to enable an understanding of the fundamentals of the method. The PO method can be broken into two steps: (1) the calculation of induced reflector surface currents, and (2) the integration of these currents (with an appropriate free space vector Green's function kernel) to determine the far-field patterns.

First, consider the calculation of the reflector surface currents. It is assumed that the field from the feed that is incident on the reflector has a spherical wavefront with amplitude tapering defined by the feed pattern. So, as a first step, the feed is mathematically modeled to determine the incident field amplitude and phase at the reflector surface. Different models are applied depending on the choice of feed used in the design; examples include waveguide feed horns, microstrip patches, dipoles, etc. Sometimes the feed model will instead employ measured feed patterns if such data is available. All models must be normalized to some prescribed radiated power level, e.g., 1 watt. Figure 12.33 shows a typical waveguide feed horn model and its associated local coordinate system.

Based on an appropriate application of equivalence principle and the induction theorem,[42–44] the currents induced on the reflector can be determined from the feed

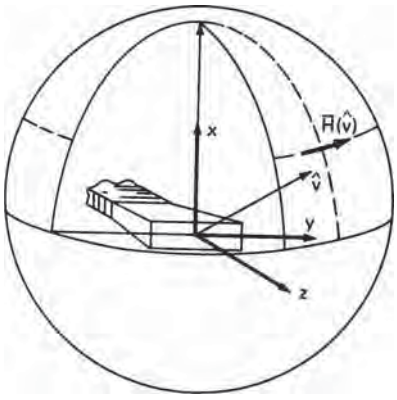


FIGURE 12.33 Waveguide feed horn model and coordinate system

$\bar{H}$-field incident at the surface of the reflector. The key premise of the equivalence principle is that fields from a scatterer (e.g., a reflector) can be represented by an "equivalent" electric current $\bar{J}$ and magnetic current $\bar{M}$ that are directly related to the incident $\bar{E}$ and $\bar{H}$ fields via

$$\bar{J} = \hat{n} \times \bar{H} \tag{12.31}$$

$$\bar{M} = -\hat{n} \times \bar{E} \tag{12.32}$$

where $\hat{n}$ is the normal to the reflector surface. The application of the equivalence principle to PO-based reflector analyses as shown here is a specialized case wherein the reflector is an electric conductor, and contributions from surface currents on the backside of the reflector are deemed negligible. An appropriate application of image theory[42–44] imposes a zero tangential E-field (Eq. 12.32) and doubles the electric current, $\bar{J}$ (Eq. 12.31), resulting in an equivalent surface current represented by

$$\bar{J} = 2\hat{n} \times \bar{H} \tag{12.33}$$

Now consider a generalized reflector surface as shown in Figure 12.34. The surface is divided into rectangular grid regions of area dA that intercept the incident feed field. If the feed far-field H-field pattern is $\bar{H}(\hat{v})$, polarized in the direction $\hat{v}$, then the incident H-field at the reflector is

$$\bar{H} = \bar{H}(\hat{v})\,(\hat{v} \cdot \hat{n})\, e^{-\mathrm{j}kr} dA/4\pi r \tag{12.34}$$

Combining Eqs. 12.34 and 12.33 yields the equivalent surface current $\bar{J}$ at area dA

$$\bar{J} = 2\hat{n} \times \bar{H}(\hat{v})\,(\hat{v} \cdot \hat{n})\, e^{-\mathrm{j}kr} dA/4\pi r \tag{12.35}$$

where $\hat{n}$ is the normal to the surface, $\hat{s}$ is the observation direction (Figure 12.34), r is the distance from the feed to the reflecting surface, $k = 2\pi/\lambda$, i.e., the wavenumber, and the $e^{-\mathrm{j}kr}/4\pi r$ term accounts for the propagation phase and space loss from the feed to the reflector surface.

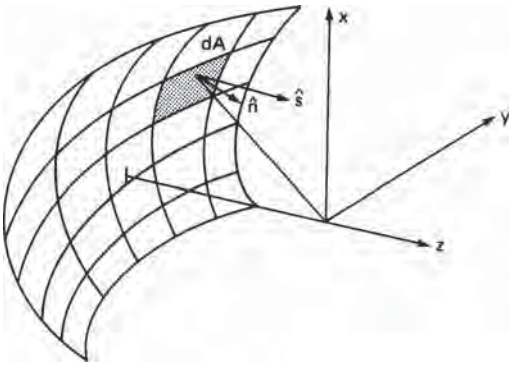


FIGURE 12.34 Generalized reflector geometry

The final step in the PO-based pattern solution is to calculate the reflector far fields via a vector integration of the product of the surface current and the free space Green's function.[44] The magnetic vector potential $\bar{A}$ is defined by the equation

$$\bar{A} = \iiint \frac{\bar{J}\, e^{-jk|r-r'|} dr'}{4\pi\, |r-r'|} \tag{12.36}$$

The vector $\bar{E}$-and $\bar{H}$-fields are related to $\bar{A}$ via simple derivatives and/or vector multiplications. In practice, Eq. 12.36 is computed as a summation of numerical integrations. Proper convergence of the field solution and resulting patterns is a function of the grid size, and this is generally achieved and verified by incrementally decreasing the grid size until the solution stabilizes. Achieving sufficient convergence in reflector PO computations within reasonable computer run times (a few minutes or less) is rarely a problem with modern computers.

Geometrical Optics (GO) Reflector Analysis, Including GTD/UTD. There are a variety of GO-based pattern reflector analysis methods, all of which are rooted in ray tracing and some of which are augmented with diffraction terms for improved accuracy. Although the simple GO method (no diffraction terms) works reasonably well, it is not generally as accurate as PO. However, two of the more common enhanced methods, Geometrical Theory of Diffraction (GTD) and Uniform Theory of Diffraction (UTD),[45] include edge diffraction and are much more accurate. UTD is essentially an enhancement of GTD wherein localized GTD singularities are corrected. The diffraction terms in UTD increase the accuracy of the basic GO solution and will properly predict the pattern asymmetries of more generalized reflector geometries. Like PO, GTD/UTD methods will generally provide high-fidelity pattern predictions for most reflector systems (center-fed, offset, single, dual, etc.) as long as the reflector size is approximately 5λ or larger. More detailed descriptions of GTD/UTD can be found in various references.[45,46]

Full-Wave Reflector Analysis Methods. Rigorous or full-wave methods include, for example, the method of moments (MOM), the finite element method (FEM), and the finite difference time domain (FDTD) method. Although these methods are very rigorous and highly accurate, they are not generally employed for reflector design/analysis because they tend to be too computationally intensive. These methods are more typically applied for precision analysis of microwave devices or electrically "smaller" antennas, e.g., radiators and feeds, that are no more than a few wavelengths in size. In recent years, hybrid methods incorporating PO or GO/GTD along with MOM, FEM, or FDTD have been developed. These methods enable enhanced reflector feed modeling, i.e., integrated with reflector analysis, and rigorous modeling of electrically small scatterers, e.g., small subreflectors or feed support struts.

Computer Codes for Reflector Design and Analysis. A number of commercial and university codes have been developed for the analysis and design of reflector systems. Two of the more well-known and commonly used codes are GRASP and the SATCOM Workbench with its NEC-REF module. GRASP is a commercially available PO-based reflector design and scattering analysis code developed by TICRA (Copenhagen, Denmark). The SATCOM Workbench was developed by The Ohio State University ElectroScience Laboratory (OSU-ESL) and is a user-friendly code that incorporates a wide variety of software modules, some of which are based upon legacy

OSU codes such as NEC-REF (OSU Reflector Code). The SATCOM Workbench is available to members of the USA Satellite Industry Code Consortium.[47]

TICRA GRASP is a generalized PO-based code intended for both reflector system design/analysis and scattering analyses. GRASP has popular GUI-versions that run on Microsoft PC-based Windows operating systems (2000, 2003, NT, XP) as well as LINUX. Although the GRASP code is PO-based, it also includes a Physical Theory of Diffraction (PTD) option as well as a Geometrical Optics (GO)/ Uniform Geometrical Theory of Diffraction (UGTD) option that can be turned on when needed. The code is very general in that it can model standard conic section-based reflectors as well as arbitrarily shaped surfaces or scatterers if desired. It also has a suite of feed models to draw upon and has tools to enable feed array modeling. Figure 12.35 shows a snapshot of the GRASP GUI window, including a rendering of an array-fed reflector and its associated beam rosetta, i.e., an overlay of beams associated with each of the individual array feeds. Some other notable features of the GRASP code include scattering models for mesh, frequency selective or lossy reflector surfaces, and a spherical wave expansion (SWE) feature for modeling of systems wherein the reflector or scatterer is in the near field of the feed source. Finally, there are various other antenna software modules available from TICRA that can be used in conjunction with GRASP. One notable example is Physical Optics Shaper (POS), an optimizer module for shaped reflector synthesis and/or feed array amplitude/phase weight synthesis. More information is available at www.ticra.com.

The OSU-ESL SATCOM Workbench[47] is a Microsoft Windows–based modular software package for PC platforms running Windows 95/98/2000/XP. It contains a variety of modules or so-called wizards in addition to the generalized GO/GTD-based reflector/scattering code backbone. The core GO/GTD reflector/scattering

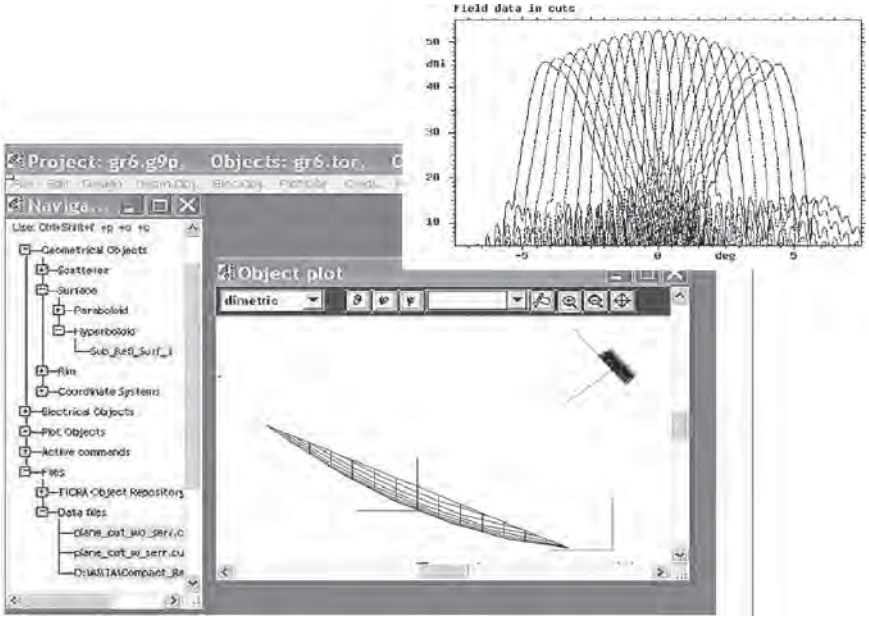


FIGURE 12.35 TICRA GRASP GUI, multibeam array-fed reflector model and patterns

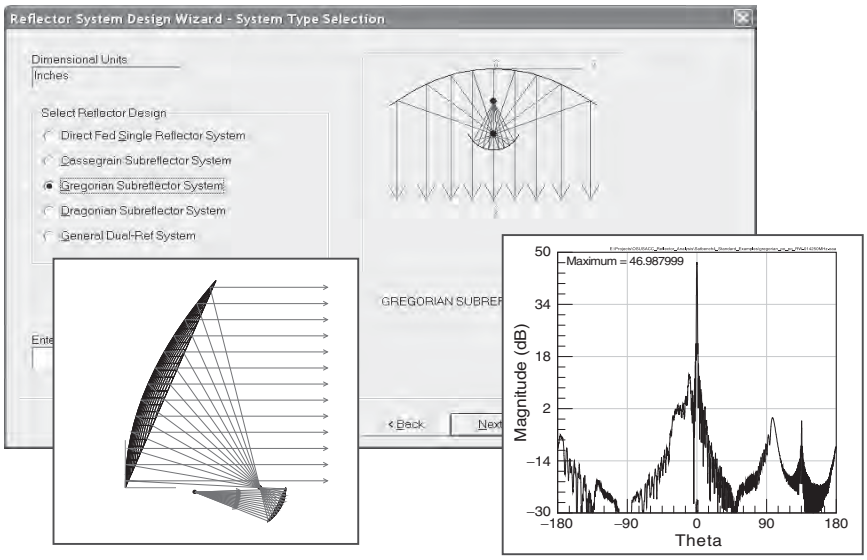


FIGURE 12.36 OSU-ESL SATCOM Workbench GUI, offset Gregorian reflector model and pattern

code portion of the SATCOM Workbench is based upon the legacy NEC-REF and NEC-BSC codes. The EM Workbench, like GRASP, is very general in its capabilities, i.e., it can handle a diversity of reflector/scatterer shapes and combinations and also has various feed models to draw upon. It also has feed and array wizards that employ full-wave EM models that can be used as sources for reflector and scattering design/analyses. Figure 12.36 shows a snapshot of the Workbench GUI window, including a rendering of an offset-fed reflector and its associated beam pattern. More information is available at esl.eng.ohio-state.edu.

12.6 MECHANICAL DESIGN CONSIDERATIONS

Reflector mechanical design is a detailed discipline unto itself with a multitude of factors to consider. Furthermore, designs vary significantly depending upon many factors, including platform, reflector size, environment, frequency of operation, scan/FOV, and cost constraints. Although it is not within the scope of this chapter to address mechanical design in detail, a brief survey of design factors and considerations provides some useful insights.

The platform, i.e., vehicle, or installation site is a significant driver for radar sensors in general, including the antenna. *Platform* is a generic term referring to the vehicle where the radar and antenna are installed. Typical radar platforms include pedestal (fixed site), ground vehicles, ships, airplanes, UAVs, and spacecraft/satellites. The following short section is devoted to platform impacts and some key associated design drivers. These include mass, volume (stowage/deployment), gimbals (precision mechanical positioning systems), materials, and mechanical tolerances. Finally, there is a brief discussion of environmental design considerations and radomes.

Installation Is a Significant Mechanical Design Driver. The platform is generally a dominant mechanical design driver because it determines the environment (thermal, vibration, etc.), and it typically drives the available size, weight, and power (SWAP) for the radar and the reflector antenna. Table 12.3 provides a qualitative comparison of typical design requirements and features for radar reflector antennas on ground-based, ship-based, airborne, and spaceborne platforms.

Mass, Volume, Stowage, Deployment, and Gimbaling. The degree to which these five factors, *mass, volume, stowage, deployment*, and *gimbaling* drive the reflector design vary in accordance with the reflector system and the platform. However,

TABLE 12.3 Mechanical Design Drivers for Reflector Antenna Systems as a Function of Platform

	PLATFORM DRIVES REFLECTOR ANTENNA MECHANICAL DESIGN			
	Ground-based	Ship-based	Airborne	Spaceborne
Mass	• Typically not a major driver. • May be a driver if field deployment is required.	• Typically not a major driver.	• Typically a significant driver. • Could be a major driver depending on size of antenna and platform.	• A major driver; launch costs are very high and are driven by available volume and mass for radar payload. • Use of light weight materials is important.
Volume	• Typically not a major driver. • May be a driver if field deployment is required	• Depends, a driver in some cases	• Typically a significant driver • Could be a major driver depending on size of antenna and platform	• A major driver; launch costs are very high and are driven by available volume and mass for radar payload. • Once in space, antenna deployments are typically needed, e.g., unfolding, etc.
Thermal	• Can be a major driver. • Radars are typically high power. • High power densities at feed or feed array are common. • Cooling system design to maintain feed temp can be a challenge.	• Can be a major driver. • Radars are typically high power. • High power densities at feed or feed array are common. • Cooling system design to maintain feed temp can be a challenge.	• Can be a major driver. • Radars are typically high power. • High power densities at feed or feed array are common. • Cooling system design to maintain feed temp can be a challenge.	• A major driver, need sophisticated modeling to predict solar heating throughout orbit. • Passive cooling systems are generally employed, e.g., heat pipes.
Vibration	• Typically not a major driver. • However, must consider vibration during transport (truck, airplane, other?).	• Typically not a major driver.	• Typically a major driver.	• Typically a major driver, driven by launch vehicle (rocket).

TABLE 12.3 Mechanical Design Drivers for Reflector Antenna Systems as a Function of Platform (*Continued*)

PLATFORM DRIVES REFLECTOR ANTENNA MECHANICAL DESIGN				
	Ground-based	Ship-based	Airborne	Spaceborne
Stowage/ Deployment	• Can be a driver if system is transportable.	• Typically not a requirement.	• Typically not a requirement, but there are exceptions.	• Typically required, generally a major design driver. • Reliability of deployment is a major concern; mission depends on it.
Other	• Is transportability a requirement? Are there multiple environments to which the radar antenna will be exposed?	• Where on the ship is the radar antenna? Is it covered with a radome? Will it be exposed to water or wave slap?	• Volume is a major constraint, not much room for antenna apertures.	• Specialized environments. • Launch drives vibration and acoustic loads. • Radiation a driver at some orbits. • Thermal extremes, including gradients, are a concern.

mass and volume constraints generally have a significant impact on the reflector system design. Furthermore, some sort of stowage and deployment of the reflector is sometimes required, especially for larger reflectors. These considerations and constraints drive the choices of materials, structural designs, passive and active mechanisms, etc. It is beyond the scope of this chapter to address this topic in detail. However, it's useful to show a couple examples for illustration.

Consider, first, a ground-based dual-reflector design with a 9-meter main reflector aperture. This reflector, shown in Figure 12.37, is used for an S-band meteorological radar application.[48] The panelized aluminum reflector is mechanically scanned via use of a gimbal (not shown). The feed, a dual-polarized waveguide horn, is also shown in Figure 12.37. The structural design of this large reflector was a significant task driven by the need to maintain low reflector surface distortion (less than 50 mils) with severe wind and gravity-loading forces and thermal gradients.

The second example is a space-based deployable reflector. The mesh reflector, shown in both stowed and deployed configurations in Figure 12.38, is an offset reflector with a 12.25 meter circular projected aperture.[49] This L-band design, developed by Northrop-Grumman Space Technologies Astro Aerospace group, has been successfully launched and deployed and is currently in use on several communication satellites.[50] A total of five reflectors of aperture diameters 9 meters, 12 meters, and 12.25 meters have been flown. Studies have addressed the potential usage of this reflector technology for various space-based radar applications, including weather sensing/monitoring (NEXRAD)[51] and planetary SAR mapping missions (lunar and Mars). Significant features of this reflector include its precise surface accuracy, high stiffness and stability, low mass, and reliable deployment. For example, for the reflector shown in Figure 12.38, an RMS surface accuracy of less than 50 mils from all error sources including in-orbit thermal gradients was achieved via prudent material choices and matching of the associated material coefficients of thermal expansion (CTEs). Pointing precision due to eclipse thermal snap has been measured in orbit at less than 0.01 degrees.[52]

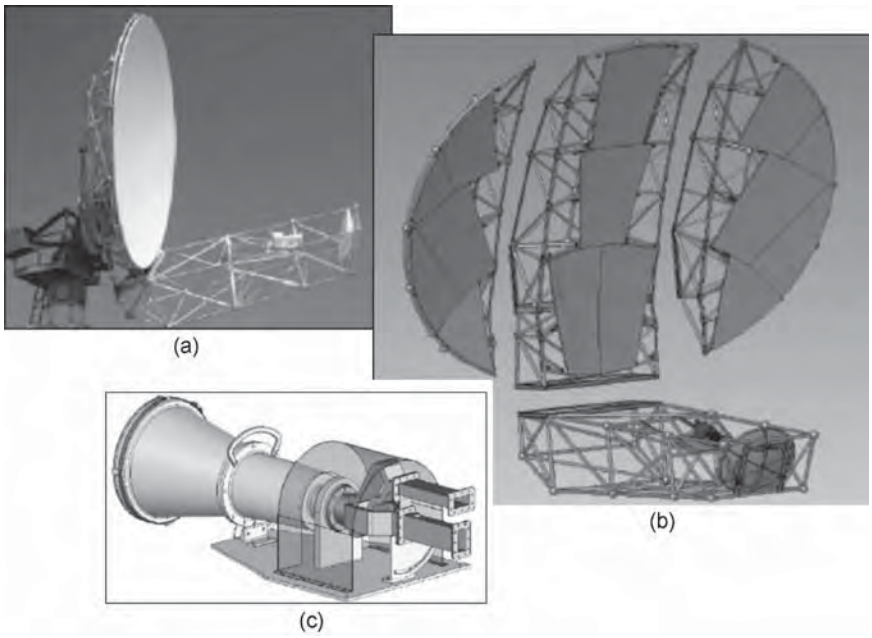


FIGURE 12.37 Ground-based S-band 9-meter dual reflector and dual-pol feed: *a*) photo of system; *b*) mechanical CAD rendering showing panelized reflector surface and support structure; and *c*) dual-polarized waveguide feedhorn (*Courtesy General Dynamics C4 Systems*)

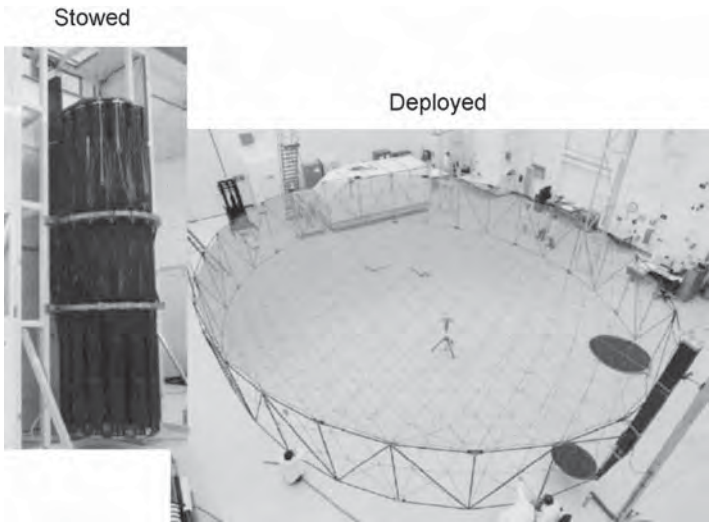


FIGURE 12.38 Space-based deployable L-band ASTROMesh reflector with 12.25 meter circular projected aperture (*Courtesy Northrop Grumman Corporation*)

Deployable space-based reflectors have also been developed by the Harris Corporation for various space-based communications applications. See www.harris.com for more details.

Because most reflector designs have at best limited electronic scan capability, a gimbal is typically needed to extend the (mechanical) FOV for the radar. The key factors or specifications that typically drive the design or procurement of the gimbal are scan rate, slew requirements, acceleration/deceleration requirements, torque and load (reflector mass) requirements, power requirements, etc. It is important for the radar system engineer to understand these factors and associated practical gimbal design limits.

Environmental Factors and Considerations. The impact of environmental factors varies widely, but some of the factors that are typically significant include thermal, vibration, and exposure (e.g., salt, sand, water, radiation, etc.). Thermal effects, i.e., spatial or temporal temperature variations, are of particular concern for spaceborne sensors that often see very significant temperature changes (greater than 100° C temporal or spatial variation is not uncommon). Vibration is another concern, especially for airborne and spaceborne reflectors. Reflectors on these platforms generally see particularly stressing vibration levels driven by airplane or launch vehicle (rocket) environments. Potential exposure to salt, sand, water, etc., is highly dependent upon platform, use of radome (or not), etc., but must be considered in the design.

Radomes. Radomes are used when it is necessary to protect antennas from adverse environmental effects. Ideally, a radome should be perfectly transparent to the RF radiation from (or to) the antenna and yet be able to withstand such environmental effects as wind, rain, hail, snow, ice, sand, salt spray, lightning, and (in the case of high-speed airborne applications) thermal, erosion, and other aerodynamic effects. In practice, these environmental factors determine the mechanical design of the radome, and the desire for ideal RF transparency must be compromised because mechanical and electrical requirements are often in conflict.

Radomes cause four major electrical effects on antenna performance. *Beam deflection* is the shift of the electrical axis, which is critical for tracking radar. *Transmission loss* is the measure of energy lost by reflection and absorption. The *reflected power* causes antenna mismatch in small radomes and sidelobes in larger radomes.

Radome design is a specialized art, and many books[53,54] are devoted to its intricate details. This section makes no attempt to provide radome design information as such but instead is aimed at making the radar system designer aware of the basic concepts behind the types of radomes available for various applications.

Three general classes of radome are of interest for reflector antenna applications: *feed covers*, which often have to endure pressure, high voltage, and heating; *covers attached to the reflector*, which alter the pattern in a fixed manner; and *external radomes*, within which the antenna moves. External radomes are the most common and will, therefore, be emphasized. Within each class, a variety of skin and skin-supporting designs is available to minimize the electrical effects under the constraints of the environment. The radome skin may be rigid, supported by a framework, or air-supported.

The most common rigid radome-wall structures are shown in Figure 12.39 and are known as homogeneous single layer, A-sandwich, B-sandwich, C-sandwich, multiple-layer sandwich, and dielectrics with metal inclusions.

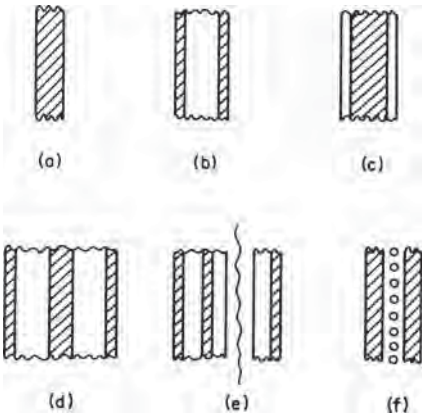


FIGURE 12.39[54] Common radome-wall cross sections: (*a*) single layer, (*b*) A-sandwich, (*c*) B-sandwich, (*d*) C-sandwich, (*e*) multiple-layer sandwich, and (*f*) dielectrics with metal inclusions

Single Layer. The homogeneous single-layer radome has been used in many radome applications. Materials for this type have included fiberglass-reinforced plastics, ceramics, elastomers, and monolithic foam. The optimum thickness for a single layer is a multiple of a half wavelength in the dielectric material at the appropriate incidence angle, but many single-layer radomes are simply thin-wall approximations to the zero-thickness case.

A-Sandwich. A commonly used radome-wall cross section is the A-sandwich, which consists of two relatively dense thin skins and a thicker low-density core. This configuration exhibits high strength-to-weight ratios. The skins are generally fiberglass-reinforced plastics, and the core is a foam or honeycomb. Inorganic skin and core sandwiches also have been developed for high-temperature applications. As a rule, the skins of the sandwich are made symmetrical or of equal thickness to allow midband cancellation of reflections.

B-Sandwich. In contrast to the A-sandwich, the B-sandwich is a three-layer configuration whose skins have a dielectric constant lower than that of the core material. This wall cross section is heavier than that of the A-sandwich because of the relatively dense core. The B-sandwich is not commonly used because the core dielectric constant is quite high for a proper match.

C-Sandwich. The C-sandwich is a five-layer design consisting of outer skins, a center skin, and two intermediate cores. The symmetrical C-sandwich can be thought of as two back-to-back A-sandwiches. This configuration is used when the ordinary A-sandwich will not provide sufficient strength, or for certain electrical performance characteristics, or when one layer is to serve as a warm-air duct for deicing.

Multiple-Layer Sandwich. Multiple-layer sandwiches of 7, 9, 11, or more layers are sometimes considered when great strength, good electrical performance, and lightweight are required. Some of these designs have used thin layers of fiberglass

laminates and low-density cores to attain high transmission performance over large frequency bands.

Dielectric Layers with Metal Inclusions. Metal inclusions have been considered for use with dielectric layers to achieve frequency filtering, broad-frequency-band performance, or reduced-thickness radomes. Thin layers of metal inclusions exhibit the characteristics of lumped circuit elements shunted across a transmission line. For example, a grid of parallel metal wires exhibits the properties of a shunt-inductive susceptance.

There are many additional radome design issues and application specific considerations and design drivers, but these are beyond the scope of this chapter.

ACKNOWLEDGMENTS

The authors wish to acknowledge and thank Helmut Schrank, Gary Evans, and Daniel Davis (also co-author of this chapter) for Chapter 6, "Reflector Antennas," in the second (1990) edition of this handbook.[55] We are grateful for their contributions and portions of the second edition chapter have been retained.

Portions of the material in the "Radome" subsection were adapted from Chapter 14,"Radomes," in the first (1970) edition of this handbook,[56] which was authored by Vincent J. DiCaudo.

REFERENCES

1. J. D. Kraus, *Antennas*, 2nd Ed., New York: McGraw-Hill Book Company, 1988: Sec. 2-34.
2. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, Chapter 8, New York: John Wiley and Sons, 1981.
3. C. M. Knop, "On the front to back ratio of a parabolic dish antenna," *IEEE Trans. Antennas Propag*., vol 24, pp. 109–111, January 1976.
4. W. V. T. Rusch, "Scattering from a hyperboloidal reflector in a cassegrain feed system," *IEEE Trans*., vol. AP-11, pp. 414–421, July 1963.
5. C. L. Gray, "Estimating the effect of feed support member blocking on antenna gain and sidelobe level," *Microwave J*., pp. 88–91, March 1964.
6. *Microwave Engineers Handbook and Buyers Guide*, New York: Horizon House, 1964, p. 143.
7. W. W. Mumford, "Some technical aspects of microwave radiation hazards," *Proc. IRE*, pp. 427–447, February 1961.
8. J. Ruze, "The effect of aperture errors on the antenna radiation pattern," *Nuovo Cimento, Suppl*., vol. 9, no. 3, pp. 364–380, 1952.
9. J. Ruze, "Antenna tolerance theory—A review," *Proc. IEEE*, vol. 54, pp. 633–640, April 1966.
10. S. Silver (ed.), *Microwave Antenna Theory and Design*, MIT Radiation Laboratory Series, vol. 12, New York: McGraw-Hill Book Company, 1949.
11. Y. T. Lo, "On the Beam Deviation Factor of a Parabolic Reflector," *IRE Trans*., vol. AP-8, pp. 347–349, May 1960.
12. P. D. Potter, "Application of spherical wave theory to Cassegrainian-fed paraboloids," *IEEE Trans*., vol. AP-15, pp. 727–736, November 1967.
13. R. C. Johnson and H. Jasik (eds.), *Antenna Engineering Handbook*, 2nd Ed., New York: McGraw-Hill Book Company, 1984, pp. 32–11, 32–12.

14. Y. T. Lo and S. W. Lee (eds.), *Antenna Handbook: Theory, Applications and Design, Reflector Antennas*, Chapter 15, New York: Van Nostrand Reinhold Co. Inc., 1988.
15. C. J. Sletten, "The theory of reflector antennas," Air Force Cambridge Res. Lab., AFCRL-66-761, Phys. Sci. Res., Paper 290, 1966.
16. K. S. Kelleher and H. P. Coleman, "Off-axis characteristics of the paraboloidal reflector," Naval Res. Lab. Rept. 4088, 1952.
17. A. W. Rudge and N. A. Adatia, "Offset parabolic reflector antennas: A review," *Proceedings IEEE*, vol. 66, no. 12, pp. 1592–1618, December 1978.
18. D. G. Kielsy, "Parabolic cylinder aerials," *Wireless Eng*., vol. 28, pp. 73–78, March 1951.
19. R. L. Fante et al., "A parabolic cylinder antenna with very low sidelobes," *IEEE Trans*., vol. AP-28, pp. 53–59, January 1980.
20. P. W. Hannan, "Microwave antennas derived from the cassegrain telescope," *IRE Trans*., vol. AP-9, pp. 140–153, March 1961.
21. P. D. Potter, "Aperture illumination and gain of a Cassegrainian system," *IEEE Trans*., vol. AP-11, pp. 373–375, May 1963.
22. W. V. T. Rusch, "Scattering from a hyperboloidal reflector in a Cassegrainian feed system," *IEEE Trans*., vol. AP-11, pp. 414–421, July 1963.
23. E. J. Wilkinson and A. J. Applebaum, "Cassegrain systems," *IRE Trans*., vol. AP-9, pp. 119–120, January 1961.
24. C. J. Sletten et al., "Offset dual reflector antennas for very low sidelobes," *Microwave J*., pp. 221–240, May 1986.
25. W. V. Rusch, "The current state of the reflector antenna art," *IEEE Trans. Antennas Propag*., vol. AP-32, no. 34, pp. 319–320, April 1984.
26. T. Haeger and J. J. Lee, "Comparisons between a shaped and nonshaped small cassegrain antenna," *IEEE Trans. Antennas Propag*., vol. 38, no. 12, December 1990.
27. R. A. Pearson, E. Elshirbini, and M. S. Smith, "Electronic beam scanning using an array-fed dual offset reflector antenna," *IEEE AP-S Int. Symp. Dig*., pp. 263–266, June 1986.
28. E. P. Ekelman and B. S. Lee, "An array-fed, dual-reflector antenna system (of offset confocal paraboloids) for satellite antenna applications," *IEEE Symp. Antennas Propag*., pp. 1586–1589, 1989.
29. H. K. Schuman and D. R. Pflug, "A phased array feed, dual offset reflector antenna for testing array compensation techniques," *IEEE Symp. Antennas Propag*., pp. 466–469, 1990.
30. W. D. Fitzgerald, "Limited electronic scanning with a near-field Cassegrainian system," Technical Report 484, MIT Lincoln Laboratory, 24 September 1971.
31. T. Lee, "A study of spherical reflectors as wide angle scanning antennas," *IEEE Trans. Antennas Propag*., vol. 7, pp. 223–226, July 1959.
32. T. Chu and P. P. Iannone, "Radiation properties of a parabolic torus reflector," *IEEE Trans. Antennas Propag*., vol. 37, no. 7, July 1989.
33. M. Skolnik, "A long range radar warning system for the detection of intercontinental ballistic missiles," MIT Lincoln Laboratory TR 128, August 15, 1956.
34. M. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2002, pp. 662, 663.
35. C. A. Balanis, *Antenna Theory Analysis and Design*, Chapters 13 and 15, New York: John Wiley and Sons, 1982.
36. A. W. Love (ed.), *Electromagnetic Horn Antennas*, New York: IEEE Press, 1976.
37. W. Cohen and C. M. Steinmetz, "Amplitude and phase sensing monopulse system parameters," *Microwave J*., pp. 27–33, October 1959.
38. D. R. Rhodes, *Introduction to Monopulse*, New York: McGraw-Hill Book Company, 1959.
39. L. J. Ricardi and L. Niro, "Design of a twelve-horn monopulse feed," *IRE Int. Conv. Rec., part. 1*, March 1961, pp. 49–56.
40. P. W. Hannan and P. A. Loth, "A monopulse antenna having independent optimization of the sum and difference modes," *IRE Int. Conv. Rec., part. 1*, March 1961, pp. 57–60.

41. B. Saka and E. Yazgan, "Pattern optimization of a reflector antenna with planar-array feeds and cluster feeds," *IEEE Trans. Antennas Propagat*., vol. 45, no. 1, January 1997.
42. R. F. Harrington, *Time Harmonic Electromagnetic Fields*, New York: McGraw-Hill, pp. 106–116, 1961.
43. L. Diaz and T. Milligan, *Antenna Engineering Using Physical Optics: Practical CAD Techniques and Software*, Boston: Artech House, 1996, pp. 193–196.
44. C. A. Balanis, "Green's functions" in *Advanced Engineering Electromagnetics*, Chapter 14, New York: John Wiley and Sons, 1989.
45. P. H. Pathak, "High frequency techniques for antenna analysis," *Proc. of the IEEE*, vol. 80, no. 1, January 1982.
46. Y. T. Lo and S. W. Lee (eds.), *Antenna Handbook: Theory, Applications and Design, Techniques for High Frequency Problems*, Chapter 4, New York: Van Nostrand Reinhold Co. Inc., 1988.
47. G. F. Paynter, T. H. Lee, and W. D. Burnside, "Expansion of existing EM Workbench for multiple computational electromagnetics codes," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, June 2003.
48. D. Brunkow, V. N. Bringi, P. C. Kennedy, S. A. Rutledge, V. Chandrasekar, E. A. Mueller, and R. K. Bowie, "A description of the CSU-CHILL National Radar Facility,". *J. Atmos. Ocean. Tech*., 17, pp. 1596–1608, 2000.
49. M. Thomson, "The astromesh deployable reflector," *IEEE Symp. Antennas and Propag*., pp. 1516–1519, 1999.
50. M. Thomson, "Astromesh Deployable Reflectors for Ku and Ka-Band Satellites," *AIAA Symp*., 2002, pp. 1–4.
51. J. K. H. Lin, H. Fang, E. Im, and U. O. Quijano, "Concept study of a 35m spherical reflector system for NEXRAD in space application," presented at 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, RI, May 1–4, 2006.
52. R. Fowell and Wang, H., "Precision pointing of the Thuraya satellite." presented at 26th AAS Guidance and Control Conference, Breckenridge, CO, February 5–9, 2003.
53. R. C. Hansen, *Microwave Scanning Antennas*, New York Academic Press, New York, 1966; Los Altos, CA: Peninsula Publishing, 1985.
54. J. D. Walton, Jr. (ed.): *Radome Engineering Handbook*, New York: Marcel Dekker, 1970.
55. M. Skolnik (ed.): *Radar Handbook*, 2nd Ed., New York: McGraw-Hill, 1990.
56. M. Skolnik (ed.): *Radar Handbook*, 1st Ed., New York: McGraw-Hill, 1970.

Chapter 13
Phased Array Radar Antennas

Joe Frank

Johns Hopkins University Applied Physics Laboratory

John D. Richards

Johns Hopkins University Applied Physics Laboratory

13.1 INTRODUCTION*

Phased Array Radars. Early radar systems used antenna arrays formed by the combination of individual radiators. Such antennas date back to the turn of the twentieth century.[1,2,3] Antenna characteristics are determined by the geometric position of the radiators and the amplitude and phase of their excitation. As radars progressed to shorter wavelengths, arrays were displaced by simpler antennas such as parabolic reflectors. For modern radar applications, the advent of electronically controlled phase shifters, switches, and transmit/receive modules has once more directed attention to array antennas. The aperture excitation may now be modulated by controlling the phase of the individual elements to give beams that are scanned electronically. The dramatic advantage of electronically steered phased arrays as compared to reflectors is provided by the time required to steer beams and the flexibility in steering. While prior radars took seconds to steer to a new location, phased arrays take microseconds. In addition, the new location can be anywhere in a hemisphere. This chapter will be devoted to arrays of this type.

Multifunction Radar. The capability of rapidly and accurately switching beams permits multiple radar functions to be performed, interlaced in time. An electronically steered array radar may track a great multiplicity of targets, illuminate a number of targets with RF energy and guide missiles toward them, and perform complete hemispherical search with automatic target selection and handover to tracking. It may even act as a communication system, directing high-gain beams toward distant receivers and transmitters. Complete flexibility is possible; search and track rates may be adjusted to best meet particular situations, all within the limitations set by the total use of time.

* Portions of this chapter that appeared in earlier editions of the *Radar Handbook* were written by the late Theodore C. Cheston, a pioneer in phased array antennas.

The antenna beamwidth may be changed to search some areas more rapidly with less gain. Frequency agility is possible with the frequency of transmission changing at will from pulse to pulse or, with coding, within a pulse. Very high powers may be generated from a multiplicity of amplifiers distributed across the aperture. Electronically controlled array antennas can give radars the flexibility needed to perform all the various functions in a way best suited for the specific task at hand. The functions may be programmed adaptively to the limit of one's capability to exercise effective automatic management and control.

Phased array theory was studied intensively in the 1960s. Technology advanced and led to a series of operational systems in the 1980s; many publications became available.[4–15] In terms of performance improvement, ultralow sidelobes (less than −40 dB) were demonstrated first in the 1970s by Westinghouse Electric Corporation's AWACS (Airborne Warning and Control System) and brought about tight tolerances in construction and phase settings. The advent of more and better computer modeling and sophisticated test equipment such as network analyzers has led to improved methods of designing well-matched apertures. Better components such as radiating elements, phase shifters, and power dividers are now available. More economical solid-state devices and memory chips have led to precision aperture phase control with corrections for frequency and temperature variations. Solid-state microwave devices hold great promise for future systems where a solid-state module is associated with each radiating element; improvements in terms of aperture control, reliability, and efficiency continue. Phased arrays can be controlled adaptively, particularly for sidelobe cancellation. This is an area where theory and understanding have advanced much. Also great progress has been made with indoor near-field antenna ranges,[16] where computer-controlled precision two-dimensional radiation patterns are derived at multiple frequencies and with scanning.

Phased arrays are very expensive. As technology advances, costs are expected to be reduced. At the same time, the quest for better performance with lower sidelobes and wider bandwidth keeps the costs high.

Phased Array Antennas. The phased array antenna has an aperture that is assembled from a great many similar radiating elements, such as slots, dipoles, or patches, each element being individually controlled in phase and amplitude. Accurately predictable radiation patterns and beam-pointing directions can be achieved.

The general planar array characteristics are readily obtained from a few simple equations that are given here but discussed later in greater detail. With the elements spaced by $\lambda/2$ (λ = wavelength) to avoid the generation of multiple beams (grating lobes), the number of radiating elements N for a pencil beam is approximately related to the beamwidth by

$$N \approx \frac{10{,}000}{(\theta_B)^2}$$

$$\theta_B \approx \frac{100}{\sqrt{N}}$$

where θ_B is the 3-dB beamwidth in degrees. The corresponding antenna gain, when the beam points broadside to the aperture, is

$$G_0 \approx \pi N \eta \approx \pi N \eta_L \eta_a$$

where η accounts for antenna losses (η_L) and reduction in gain due to unequal weighting of the elements with a nonuniform amplitude (η_a). When scanning to an angle is θ_0, the gain of a planar array is reduced to that of the projected aperture:

$$G(\theta_0) \approx \pi N \eta \cos \theta_0$$

Similarly, the scanned beamwidth is increased from the broadside beamwidth (except in the vicinity of endfire, $\theta_0 = 90°$):

$$\theta_B \text{ (scanned)} \approx \frac{\theta_B \text{(broadside)}}{\cos \theta_0}$$

The total number of beams M (with broadside beamwidth and square stacking) that fit into a sphere is approximately equal to the gain and with $\eta \approx 1$ is thus simply related to N by

$$M \approx \pi N$$

An array where the elements are fed in parallel (see Section 13.8) and scanned by phase shift, modulo 2π, has limited bandwidth; for wideband operation, constant path lengths rather than constant phases are required. The limit is given by

$$\text{Bandwidth (\%)} \approx \text{beamwidth (°)}$$

This is equivalent to limitations given by

$$\text{Pulse length} = 2 \times \text{aperture size}$$

With these criteria, the scanned radiation pattern at 60° is steered by ± one-fourth of the beamwidth at the angle of scan as the frequency is changed over the band. If all the frequencies in the band are used with equal weighting, then twice the bandwidth (half the pulse length) becomes acceptable. At a scan angle θ_0, the beam steers with frequency through an angle $\delta\theta$ so that

$$\delta\theta \approx \frac{\delta f}{f} \tan \theta_0 \qquad \text{(rad)}$$

For wider bandwidths, time-delay networks have to be introduced to supplement the phase shifters.

Conformal Arrays.[17,18] Phased arrays may conform to curved surfaces as required, for example, for flush-mounting on aircraft or missiles. If the surface has a large radius of curvature so that all the radiating elements point in substantially the same direction, then the characteristics are similar to those of a planar array even though the exact 3D position of the element has to be taken into account to calculate the required phase. A small radius of curvature is found with cylindrical (or spherical) arrays used for 360° coverage. Elements are switched to avoid sections of the antenna where they point away from the desired beam direction. Difficulties may be encountered in matching the radiating elements and in maintaining polarization purity. The discussion in this chapter will concentrate on planar phased arrays, rather than conformal arrays.

3D Volumetric Search. Three-dimensional (3D) volumetric radar search is possible with electronic scanning in both azimuth and elevation; important regions (e.g., the horizon) may be emphasized at will and searched more frequently. The radar may operate with a higher than normal false-alarm rate since targets can easily be confirmed by additional dwells. Phase control allows beams to be widened, for example, to reduce search time for the more elevated regions, where reduced ranges need less antenna gain. A separate rotating surveillance radar system may be added for extra coverage (at a second frequency) and to allow the 3D radar more time for tracking.

Monopulse Track. Phased array radars are well suited for monopulse tracking. The radiating elements of the array can be combined in three different ways to give the sum pattern and the azimuth and elevation difference patterns. Contradictory requirements in optimum amplitude distribution for sum and difference patterns exist,[19] but, as with other antenna systems, they may be independently satisfied. The sum and difference patterns are scanned simultaneously.

The difference-pattern null in a phased array system gives good beam-pointing accuracy. Absolute beam-pointing accuracies to within less than one-fiftieth of a (scanned) beamwidth have been measured with scans up to 60°.[20] The accuracy is limited by phase and amplitude errors. Since phase shift rather than time delay is used, as the frequency is changed, the direction of the null of the scanned beam is also changed, and the beam moves toward broadside with an increase in frequency.

Shaped Beams. The radiation pattern of an array may be shaped by modifying the aperture distribution. Good pattern approximations can be obtained by using phase only. In particular, the beam may be broadened by applying a spherical phase distribution to the aperture or by approximating it with a gable (triangular) phase distribution. Beams of this type are of particular interest because they are easily generated. They may be used for transmission in a system where the receiving antenna has a cluster of simultaneous beams, or, as previously discussed, they may be used in a search system to reduce the number of angular cells in regions of shorter range.

Monitoring. Electronically scanned arrays are composed of very many parts and include electronic circuitry to drive the phase shifters or switches that steer the beam. The overall reliability of such arrays can be great; graceful degradation has been claimed because the failure of as much as 10% of the components leads to a loss in gain of only 1 dB. There is, however, a degradation of (low) sidelobes. Nevertheless, the functioning of the antenna is complex, and there is need for providing test or monitoring circuitry. The decision to point a beam in a certain direction is made somewhere in the radar control system and is normally defined by two direction cosines. A test or monitoring circuit should establish the correct functioning of all components, including all beam-pointing computations, electronic drivers and phase shifters or switches, and all their interconnections. Frequent indications that the antenna system is functioning or is capable of functioning should be available. In one possible method, the phase shifters are programmed to focus on a nearby monitor probe and scan past it.[21] This will yield a close approximation of the complete radiation pattern, where gain and sidelobes can be measured and compared with previous results. The combination of individual elements and their phase shifters (and drivers) can also be checked with this configuration. The phase at each element is sequentially rotated at some low frequency; the amplitude and phase of this modulation as received by the probe relate directly to both the relative amplitude excitation of the element and its

relative phase setting.[22] Other methods have been proposed[23] where measurements are compared with previously recorded ones.

Deployment of Apertures. With planar arrays, scanning is limited by the loss in gain and the increase in beamwidth corresponding to the reduction of the aperture to its projected area. Practical extreme values of scanning are therefore in the region of 60 to 70°. A minimum of three planar array apertures is then necessary for hemispherical coverage. The antennas may be positioned as shown in Figure 13.1, permitting a view that is unimpeded by the central superstructure. The apertures would normally be tilted back from the vertical to balance the scan angles.

Radiating Elements. The most commonly used radiators for phased arrays are dipoles, slots, open-ended waveguides (or small horns), and printed-circuit "patches" (originally called *Collings radiator* after their inventor[24]). The element has to be small enough to fit in the array geometry, thereby limiting the element to an area of a little more than $\lambda^2/4$. In addition, many radiators are required, and the radiating element should be inexpensive and reliable and have identical characteristics from unit to unit.

Because the impedance and pattern of a radiator in an array are determined predominantly by the array geometry (Section 13.4), the radiating element may be chosen to suit the feed system and the physical requirements of the antenna. For example, if the radiator is fed from a stripline phase shifter, a stripline dipole would be a logical choice. If a waveguide phase shifter is used, an open-ended waveguide or a slot might be convenient. At the lower frequencies, where coaxial components are prevalent, dipoles have been favored. A ground plane is usually placed about $\lambda/4$ behind an array of parallel dipoles so that the antenna forms a beam in only one hemisphere.

For limited scanning (say, less than 10°), it is possible to use directive radiators having dimensions of height and width of several wavelengths. With such separation, the mutual coupling effects (see Section 13.4) can be small, and the pattern and impedance of an element in the array approach those of the isolated element.

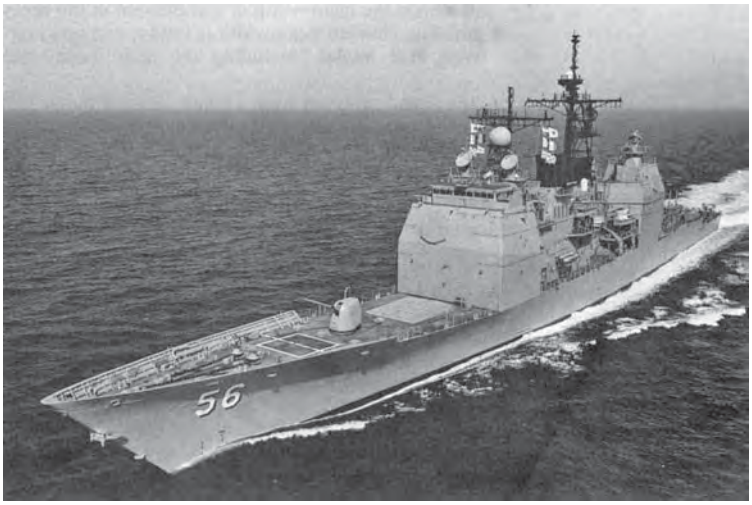


FIGURE 13.1 Guided missile cruiser showing two out of four phased array antennas (*Courtesy of Ingalls Shipbuilding Division of Litton*)

The element must be chosen to give the desired polarization, usually vertical or horizontal. The special case of circular polarization is discussed below.

If polarization diversity is required or if an array is required to transmit one polarization and receive the orthogonal or both polarizations, either crossed dipoles or circular or square radiators seem suitable. With appropriate feed systems, both are capable of providing vertical and horizontal polarization independently and may be combined to provide any desired polarization, including circular. Such polarization diversity adds considerable complexity, requiring two feed systems or switches at the radiating element level.

Circular Polarization. From the point of view of the antenna designer, circular polarization is possible, though difficulties may be encountered in matching for large scan angles. On scanning, a component of the undesired orthogonal polarization will be generated,[25] and some provision should be made to absorb that energy.[26] With a conventional circularly polarized antenna, such as a parabolic dish with a circularly polarized feed, good circularity may be obtained over part of the main beam, with rapid deterioration over the rest of the pattern. With a planar array, the relevant beamwidth is the beamwidth of the element in the array rather than the array beamwidth.

With circular polarization, the signal returned from a single-bounce target (i.e., a sphere or flat plate) will require an antenna matched to the opposite sense of circular polarization from that transmitted. If the same antenna is used, then single-bounce targets are rejected. Such a system can, therefore, give a measure of suppression of rain echoes,[27] ideally amounting to

$$20 \log (e^2 + 1)/(e^2 - 1) \quad \text{dB}$$

where e is the voltage-ellipticity ratio. An early model of a Raytheon reflectarray gave an ellipticity ratio of less than 1.5 dB with scans up to 30°, corresponding to a theoretical rain rejection of at least 15 dB. At the same time, an aircraft target would typically lose approximately 3 dB, leaving a relative net improvement of 12 dB of rain rejection.

Phased Arrays with Very Wide Bandwidth. A radar system that has the capability of changing frequency over a very wide bandwidth can, with advantage, adapt its transmission to take into account frequency-dependent multipath characteristics, target response, environmental conditions, interference, and jamming. Further, wideband processing can give fine range resolution.

Phased arrays have the potential of operating over very wide bandwidths. The high end of the frequency band is limited by the physical size of the elements, which must be spaced close enough in the array to avoid the generation of grating lobes (see Section 13.2). For wide instantaneous bandwidth (rather than tunable bandwidth), time delays have to be added to prevent the beam from being scanned as the frequency is changed.

The impedance of the radiating element at the aperture (with closely spaced elements) is approximately independent of frequency, but the element must be matched over the wide bandwidth. This is difficult to achieve without exciting harmful surface waves when scanning. Nevertheless, matching with octave bandwidth for scanning to ±60° has been achieved.

Limited Scanning.[28] If scanning is limited to a small angular volume, considerable simplifications become possible. The total number of active controls can be reduced to

about equal the total number of beams. Subarrays (see Section 13.8) may be formed, each with only one phase control of a size such that the subarray beamwidth includes all the scan angles. Alternatively, a small phased array could be placed in the focal region of a large reflector to scan the narrow beamwidth of the reflector over a limited scan angle.

Scanning of Arrays

Phase Scanning. The beam of an antenna points in a direction that is normal to the radiated phase front. In phased arrays, this phase front is adjusted to steer the beam by individual control of the phase of excitation of each radiating element. This is indicated in Figure 13.2*a*. The phase shifters are electronically actuated to permit rapid scanning and are adjusted in phase to a value between 0 and 2π rad. With an interelement spacing s, the incremental phase shift ψ between adjacent elements for a scan angle θ_0 is $\psi = (2\pi/\lambda)s \sin \theta_0$. If the phase ψ is constant with frequency, the scan angle θ_0 is frequency-dependent.

Time-Delay Scanning. Phase scanning was seen to be frequency-sensitive; however, time-delay scanning is independent of frequency. Delay lines are used instead of phase shifters, as shown in Figure 13.2*b*, providing an incremental delay from element to element of $\tau = (s/c) \sin \theta_0$, where c = velocity of propagation. Individual time-delay circuits (Section 13.7) are normally too cumbersome to be added to each radiating element. A reasonable compromise may be reached by adding one time-delay network to a group of elements (subarray) where each element has its own phase shifter.

Frequency Scanning.[29] Frequency rather than phase may be used as the active parameter to exploit the frequency-sensitive characteristics of phase scanning. Figure 13.2*c* shows the arrangement. At one particular frequency, all radiators are in phase. As the frequency is changed, the phase across the aperture tilts linearly, and the beam is scanned. Frequency-scanning arrays are relatively simple and inexpensive

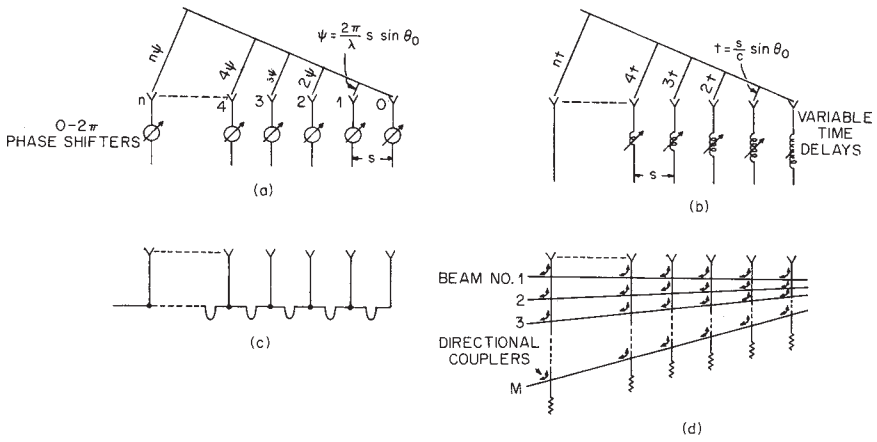


FIGURE 13.2 Generation of scanned beams: (*a*) phased array, (*b*) time-delay array, (*c*) frequency-scanned array, and (*d*) Blass-type array

to implement. They have been developed and deployed in the past to provide elevation-angle scanning in combination with mechanical horizontal rotation for 3D radars. A chapter in the first edition of this handbook was devoted to this approach, which since then has received much less attention; frequency is too important a parameter for achieving high-range resolution, electronic counter-countermeasures, and multiple radar occupancy to give it up for antenna scanning. Frequency scanning is seldom used anymore.

IF Scanning. For receiving, the output from each radiating element may be heterodyned (mixed) to an intermediate frequency (IF). All the various methods of scanning are then possible, including the beam-switching system described below, and can be carried out at IF, where amplification is readily available and lumped constant circuits may be used.

Digital Beamforming.[30–32] For receiving, the output from each radiating element may be amplified and digitized. The signal is then transferred to a computer for processing, which can include the formation of multiple simultaneous beams (formed with appropriate aperture illumination weighting) and adaptively derived nulls in the beam patterns to avoid spatial interference or jamming. Limitations are due to the availability and cost of analog-to-digital (A/D) converters and to their frequency and dynamic-range characteristics. Partial implementation is possible by digitizing at subarray levels only.

Beam Switching. With properly designed lenses or reflectors, a number of independent beams may be formed by feeds at the focal surface. Each beam has substantially the gain and beamwidth of the whole antenna. Allen[33] has shown that there are efficient equivalent transmission networks that use directional couplers and have the same collimating property. A typical form, after Blass,[34] is shown in Figure 13.2*d*. The geometry can be adjusted to provide equal path lengths, thus providing frequency-independent time-delay scanning. Another possible configuration providing multiple broadband beams uses parallel plates containing a wide-angle microwave lens.[35,36] Each port corresponds to a separate beam. The lens provides appropriate time delays to the aperture, giving frequency-invariant scanning. The beams may be selected through a switching matrix requiring $M - 1$ single-pole–double-throw (SPDT) switches to select one out of M beams. The beams are stationary in space and overlap at approximately the 4 dB points. This is in contrast to the previously discussed methods of scanning, where the beam can be steered accurately to any position. The beams all lie in one plane. The system may be combined with mechanical rotation of the antenna, giving vertical-switched scanning for 3D coverage. Much greater complexity is required for a system switching beams in both planes.

Multiple Simultaneous Beams. Instead of switching the beams, as described in the preceding paragraph, all the beams may be connected to separate receivers, giving multiple simultaneous receive beams. The transmitter radiation pattern would need to be wide to cover all the receive beams. Such multibeam systems have found application in combination with mechanical rotation for 3D coverage.

Multiple Independently Steered Beams. Independent multiple beams may be generated with a single beamformer by modifying both amplitude and phase at the aperture. This can be seen from Figure 13.3, where, for example, two independent

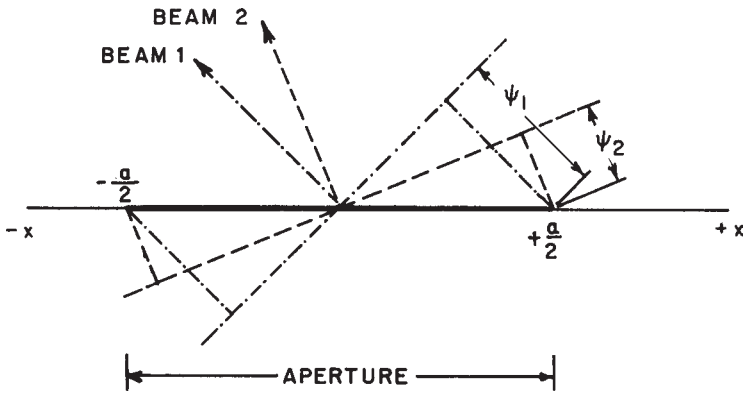


FIGURE 13.3 Aperture distribution giving two beams

beams are generated. Both beams have the same amplitude (voltage) distribution $F(x)$ but differently inclined linear phased fronts. The total aperture excitation with both beams is

$$F(x,\psi) = F(x)e^{j2\psi_1(x/a)} + F(x)e^{j2\psi_2(x/a)} = 2F(x)\cos\left[(\psi_1 - \psi_2)\frac{x}{a}\right]e^{j(\psi_1+\psi_2)(x/a)}$$

That is, the aperture amplitude distribution required for two separate beams varies cosinusoidally, and the phase distribution is linear and has the average inclination.

In most phased array systems, only the phase can be controlled. Ignoring the required amplitude variations still leads to good approximations for forming multiple beams, by superimposing the various required phase-shifter settings (modulo 2π). In the case of two beams, the aperture phase slope has the average inclination and varies periodically from 0 to π.

Vertical Scan Only. A greatly simplified phased array system becomes possible if there is no need for multifunction capabilities, including fire control, where a beam may have to be pointed in any given direction at any time. The array is scanned in the vertical plane only and mechanically rotated to give azimuth coverage. The number of phase control points is then reduced to the number of horizontal rows. In the case of a ship's surveillance radar, the antenna should be positioned as high as possible to avoid shadowing by the superstructure, but the pedestal need not be stabilized since stabilization can be achieved by electronic beam steering. Scanning can be in the form of phase scanning or multiple simultaneous beams may be used on receive with a wide antenna pattern on transmit.

13.2 ARRAY THEORY

Array with Two Elements. Figure 13.4 shows two isotropic radiators that are spaced by a distance s and excited with equal amplitude and phase. With unity input

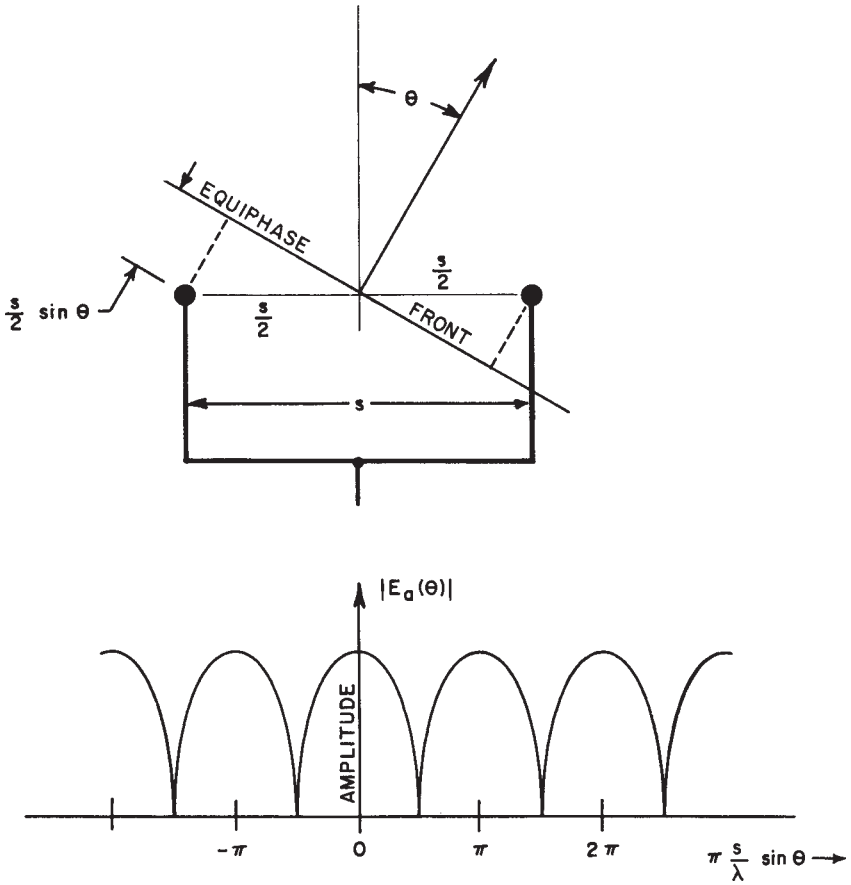


FIGURE 13.4 Radiation pattern of two isotropic radiators

power, the vector sum of their contributions, added at a great distance as a function of θ, is the radiation pattern

$$E_a(\theta) = \frac{1}{\sqrt{2}}[e^{j(2\pi/\lambda)(s/2)\sin\theta} + e^{-j(2\pi/\lambda((s/2)\sin\theta}]$$

where θ is measured from the broadside direction. Normalizing, to get unity amplitude when $\theta = 0$, and simplifying give

$$E_a(\theta) = \cos\left[\pi \frac{s}{\lambda} \sin\theta\right] \tag{13.1}$$

The absolute value of $E_a(\theta)$ is plotted in Figure 13.4 as a function of $\pi(s/\lambda)\sin\theta$. If the plot had been in terms of the angle θ, the lobes would have been found to increase in width as $|\theta|$ increased. The main beam occurs when $\sin\theta = 0$. The other lobes have the same amplitude as the main beam and are referred to as *grating lobes*.

They occur at angles given by $\sin \theta = \pm [m/(s/\lambda)]$, where m is an integer. For the half space given by $-90° < \theta < +90°$, there are $2m'$ grating lobes, where m' is the largest integer smaller than s/λ. If $s < \lambda$, grating-lobe maxima do not occur, and the value at $\pm 90°$ is cos $(\pi s/\lambda)$. This value is for isotropic radiators and is reduced if the radiators have directivity.

Linear Array.[37] With a linear array of N isotropic radiators, excited with equal amplitudes and phase and separated by distances s, as shown in Figure 13.5, the condition for the occurrence of grating lobes is unchanged from the simpler case just considered. They occur for the same values of $\pi(s/\lambda) \sin \theta$, but the width of the lobes is reduced, and they are separated by minor lobes. Summing the vector contributions from all elements, with element 0 as phase reference, gives

$$E_a(\theta) = \frac{1}{\sqrt{N}} \sum_{n=0}^{n=N-1} e^{j(2\pi/\lambda)\, ns \quad \sin \theta}$$

The factor $1/\sqrt{N}$ shows that each element is energized with $1/N$ of the (unity) input power. Normalizing the gain to unity at broadside, $\theta = 0$, gives the pattern

$$E_a(\theta) = \frac{\sin[N\pi(s/\lambda) \sin \theta]}{N \sin [\pi(s/\lambda) \sin \theta]} \tag{13.2}$$

$E_a(\theta)$ gives the radiation pattern for isotropic radiators and is known as the *array factor*. It is shown in Figure 13.6 for $N = 10$. This pattern is repetitive, and the locations of the adjacent grating lobes at angles θ_1 and θ_2 are separated by $\pi(s/\lambda)$ $(\sin \theta_1 - \sin \theta_2) = \pi$.

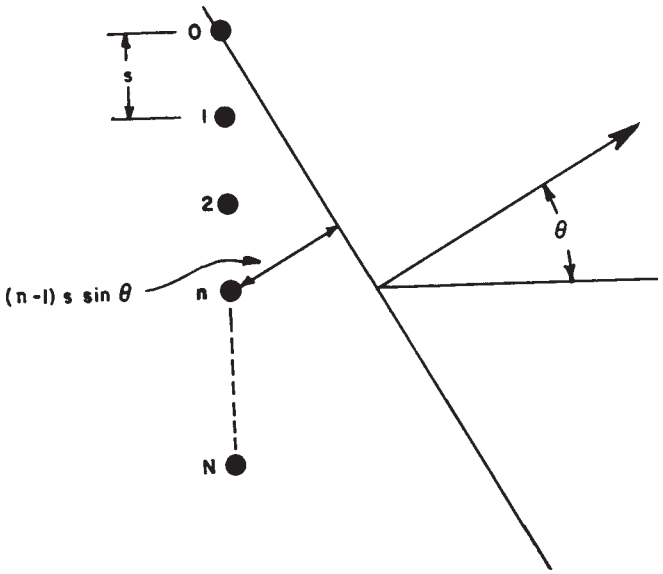


FIGURE 13.5 Linear array with N radiators uniformly spaced by a distance s

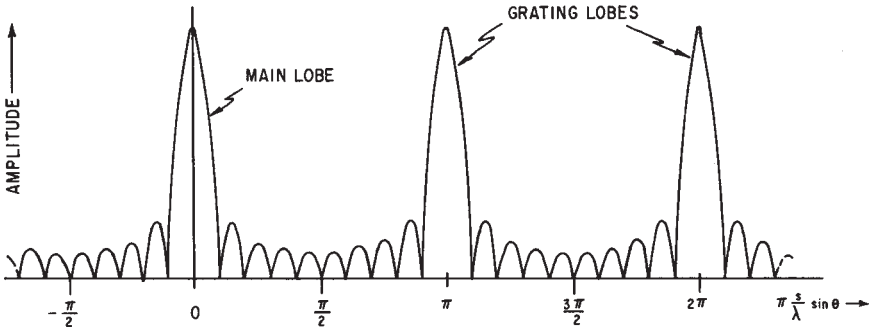


FIGURE 13.6 Array factor with $N = 10$ elements

The radiating elements are not isotropic but have a radiation pattern $E_e(\theta)$, known as the *element factor* or *element pattern*; then the complete radiation pattern $E(\theta)$ is the product of the array factor and the element pattern:

$$E(\theta) = E_e(\theta)E_a(\theta) = E_e(\theta)\frac{\sin[N\pi(s/\lambda)\sin\theta]}{N\sin[\pi(s/\lambda)\sin\theta]} \tag{13.3}$$

An approximation to the pattern of Eq. 13.2 is in the form

$$E(\theta) = \frac{\sin[\pi(a/\lambda)\sin\theta]}{\pi(a/\lambda)\sin\theta} \tag{13.4}$$

where the effective aperture is $a = Ns$, which extends by $s/2$ beyond the centers of the end elements. In contrast to the array factor, this pattern has only one maximum and is nonrepetitive. It is the well-known Fourier transform of a continuous constant-amplitude distribution. For uniform illumination, the beamwidth is

$$\theta_B = \frac{0.886}{a/\lambda}(\text{rad}) = \frac{50.8}{a/\lambda}(^\circ) \tag{13.5}$$

The first sidelobe is 13.3 dB down from the peak of the main beam.

For larger values of θ, the pattern of a continuous aperture is modified from Eq. 13.4 by the obliquity factor[38,39] 1/2 (1 + cos θ). This gives

$$E(\theta) = \tfrac{1}{2}(1+\cos\theta)\frac{\sin[\pi(a/\lambda)\sin\theta]}{\pi(a/\lambda)\sin\theta} \tag{13.6}$$

For closely spaced elements, the obliquity factor is very similar to the amplitude pattern of a well-designed (matched) radiating element, $\sqrt{\cos\theta}$, for values up to some 60 or 70°. At greater angles, the element pattern has values that are greater than those given by $\sqrt{\cos\theta}$ and that are a function of the total number of elements.[40]

Scanned Linear Arrays. The pattern of the array may be steered to an angle θ_0 by applying linearly progressive phase increments from element to element so that the phase between adjacent elements differs by $2\pi(s/\lambda)\sin\theta_0$. Equation 13.2 is then modified, giving the normalized array factor of a uniformly illuminated arrays as

$$E_a(\theta) = \frac{\sin N\pi(s/\lambda)(\sin\theta - \sin\theta_0)}{N \sin \pi(s/\lambda)(\sin\theta - \sin\theta_0)} \tag{13.7}$$

and the pattern is

$$E(\theta) = E_e(\theta)\frac{\sin N\pi(s/\lambda)(\sin\theta - \sin\theta_0)}{N \sin[\pi(s/\lambda)(\sin\theta - \sin\theta_0)]} \tag{13.8}$$

Equation 13.8 describes the fundamental response of a scanned array system. The array will have only one single major lobe, and grating-lobe maxima will not occur for $-90° < \theta < +90°$ as long as

$$\pi\frac{s}{\lambda}|\sin\theta - \sin\theta_0| < \pi$$

or

$$\frac{s}{\lambda} < \frac{1}{1+|\sin\theta_0|} \tag{13.9}$$

which is always true if $s/\lambda < ½$. When scanning is limited, the value of s/λ may be increased, for example, to $s/\lambda < 0.53$ for scanning to a maximum of 60° or $s/\lambda < 0.59$ for scanning to a maximum of ±45°.

For larger values of s/λ, grating lobes occur at angles θ_1, given by

$$\sin\theta_1 = \sin\theta_0 \pm \frac{n}{s/\lambda} \tag{13.10}$$

when n is an integer.

In the limit, the inequality (Eq. 13.9) does allow a grating-lobe peak to occur at 90° when scanning to θ_0. Even though the grating lobe is reduced when multiplied by the element pattern, it may be prudent to space the elements such that the first null of the grating lobe, rather than the peak, occurs at 90°. With N elements this more restrictive condition is given by

$$\frac{s}{\lambda} < \frac{N-1}{N} \times \frac{1}{1+|\sin\theta_0|} \tag{13.11}$$

Equation 13.8 may again be approximated by the Fourier transform of the illumination across the continuous aperture:

$$E(\theta) = ½(1+\cos\theta)\frac{\sin \pi(a/\lambda)(\sin\theta - \sin\theta_0)}{\pi(a/\lambda)(\sin\theta - \sin\theta_0)} \tag{13.12}$$

The Fourier-transform solutions for continuous apertures[19,41] may be used to approximate patterns for practical amplitude and phase distributions as long as the element-to-element spacing is small enough to suppress grating lobes.[42] Monopulse difference patterns may be approximated in the same way from the Fourier transforms of the corresponding continuous odd aperture distributions.

Element Factor and Gain of Planar Arrays. The maximum gain of a uniformly illuminated and lossless aperture of area A, with a broadside beam, is $G_{max} = 4\pi A / \lambda^2$.

With a nonuniform aperture distribution and with losses present, the gain is reduced by the efficiency term η to

$$G_{\max} = 4\pi \frac{A}{\lambda^2} \eta \tag{13.13}$$

If the aperture is considered as a matched receiver, then the amount of energy arriving from a direction θ_0 is proportional to its projected area. The gain with scanning, therefore, is

$$G(\theta_0) = 4\pi \frac{A \cos \theta_0}{\lambda^2} \eta \tag{13.14}$$

If the aperture is made up of N equal radiating elements and is matched to accept the incident power, then the contribution to the overall gain is the same from all elements, hence

$$G(\theta) = NG_e(\theta)\eta \tag{13.15}$$

where G_e is the gain per element. It follows from Eq. 13.14 that the matched element power pattern is

$$G_e(\theta) = 4\pi \frac{A}{N\lambda^2} \cos \theta \tag{13.16}$$

and the normalized radiation amplitude pattern of the (matched) element or (matched) *element pattern* is

$$E_e(\theta) = \sqrt{\cos \theta} \tag{13.17}$$

For a given element spacing s, the total number of radiators N in the area A is $N = A/s^2$, and Eq. 13.16 gives

$$G_e(\theta) = 4\pi \left[\frac{s}{\lambda}\right]^2 \cos \theta$$

When the element spacing is $s = \lambda/2$, then the power pattern of an element that is perfectly matched at all scan angles is

$$G_e(\theta) = \pi \cos \theta \tag{13.18}$$

And the peak antenna gain in the direction of scan, θ_0, is

$$G(\theta_0) = \pi N \eta \cos \theta_0 \tag{13.19}$$

where the efficiency term η accounts for losses and for a nonuniform aperture distribution. For a broadside beam $\theta_0 = 0$ and

$$G_0 = \pi N \eta \tag{13.20}$$

and the element gain is $G_e = \pi$.

Figure 13.7 shows a theoretical example of the array and element factors and the resulting pattern for a 10-element array, with element spacing $s = \lambda/2$, scanned to 60°.

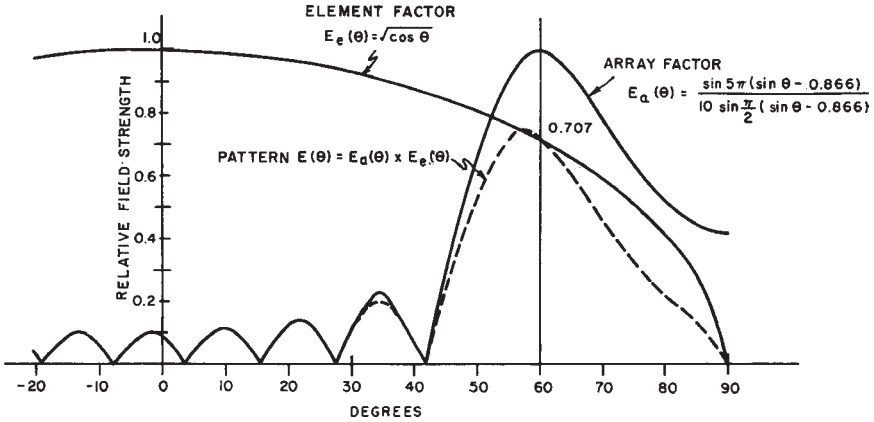


FIGURE 13.7 Ten-element linear array scanned to 60°; element spacing $s = \lambda/2$

The pattern maximum is noted to occur at less than 60° because the gain of the element pattern increases toward broadside. The element pattern value at 60° is cos 60° = 0.5 in power or 0.707 in amplitude, relative to the maximum at broadside, as expected. The sidelobes in the general region of broadside are not reduced because in that region the element pattern is approximately unity. Relative to the beam maximum, therefore, the sidelobes near broadside are increased by approximately 3 dB.

13.3 PLANAR ARRAYS AND BEAM STEERING

Planar Arrays. A planar array is capable of steering the beam in two dimensions. In a spherical-coordinate system, the two coordinates θ and ϕ define points on the surface of a unit hemisphere. As shown in Figure 13.8, θ is the angle of scan measured from broadside and ϕ is the plane of scan measured from the x axis. Von Aulock[43] has presented a simplified method for visualizing the patterns and the effect of scanning. He considers the projection of the points on a hemisphere onto a plane (Figure 13.9): the axes of the plane are the direction cosines cos α_x, cos α_y. For any direction on the hemisphere, the direction cosines are

$$\cos\alpha_x = \sin\theta\cos\phi$$
$$\cos\alpha_y = \sin\theta\sin\phi$$

The direction of scan is indicated by the direction cosines cos α_{xs}, cos α_{ys}. Here the plane of scan is defined by the angle ϕ measured counterclockwise from the cos α_x axis and is given by

$$\phi = \tan^{-1}\frac{\cos\alpha_{ys}}{\cos\alpha_{xs}}$$

The angle of scan θ is determined by the distance of the point (cos α_{xs}, cos α_{ys}) from the origin. This distance is equal to sin θ. For this reason, a representation of this sort is called sin θ space. A feature of sin θ space is that the antenna pattern shape is invariant

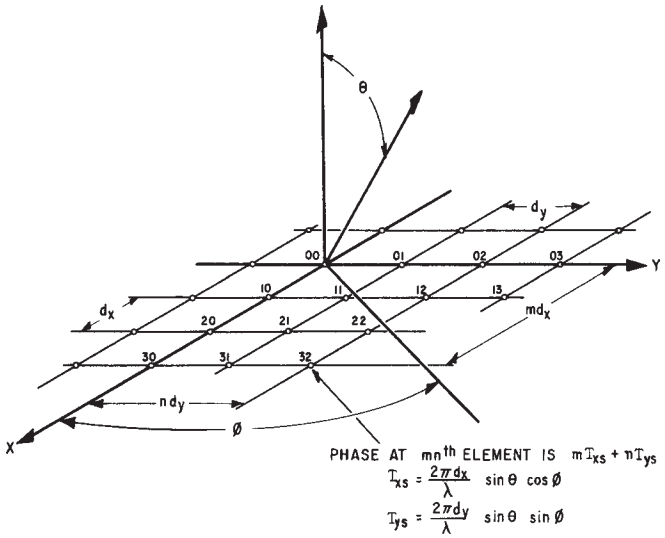


FIGURE 13.8 Planar-array-element geometry and phasing

to the direction of scan. As the beam is scanned, every point on the plot is translated in the same direction and by the same distance as is the beam maximum.

The region inside the unit circle where

$$\cos^2 \alpha_x + \cos^2 \alpha_y \leq 1$$

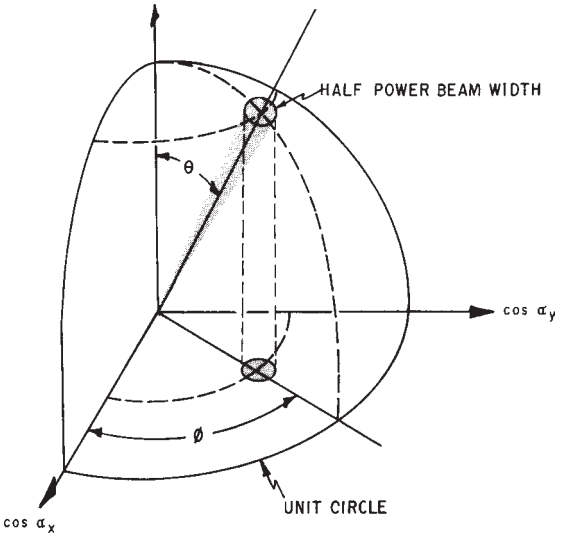


FIGURE 13.9 Projection of points on a hemisphere onto the plane of the array

is defined as *real space*, the hemisphere into which energy is radiated. The infinite region outside the unit circle is referred to as *imaginary space*. Although no power is radiated into imaginary space, the concept is useful for observing the motion of grating lobes as the array is scanned. In addition, the pattern in imaginary space represents stored energy and contributes to the element impedance in the array.

The most common element lattices have either a rectangular or a triangular grid. As shown in Figure 13.8, the *mn*th element is located at (md_x, nd_y). The triangular grid may be thought of as a rectangular grid where every other element has been omitted. The element locations can be defined by requiring that $m + n$ be even.

Calculations for the element-steering phases are greatly simplified by the adoption of the direction cosine coordinate system. In this system, the linear-phase tapers defined by the beam-steering direction ($\cos \alpha_{xs}$, $\cos \alpha_{ys}$) may be summed at each element so that the phasing at the *mn*th element is given by

$$\psi_{mn} = mT_{xs} + nT_{ys}$$

where $T_{xs} = (2\pi/\lambda)d_x \cos \alpha_{xs}$ = element-to-element phase shift in the x direction
$T_{ys} = (2\pi/\lambda)d_y \cos \alpha_{ys}$ = element-to-element phase shift in the y direction

The array factor of a two-dimensional array may be calculated by summing the vector contribution of each element in the array at each point in space. For an array scanned to a direction given by the direction cosines $\cos \alpha_{xs}$ and $\cos \alpha_{ys}$, the array factor of an $M \times N$ rectangular array of radiators may be written

$$E_a(\cos \alpha_{xs}, \cos \alpha_{ys}) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |A_{mn}| e^{j[m(T_x - T_{xs}) + n(T_y - T_{ys})]}$$

where $T_x = (2\pi/\lambda)\, d_x \cos \alpha_x$
$T_y = (2\pi/\lambda)\, d_y \cos \alpha_y$
A_{mn} = amplitude of *mn*th element

An array may be visualized as having an infinite number of grating lobes only one of which (namely, the main beam) is desired in real space. It is convenient to plot the position of the grating lobes when the beam is phased for broadside and observe the motion of these lobes as the beam is scanned. Figure 13.10 shows the grating-lobe locations for both rectangular and triangular spacing. For a rectangular array, the grating lobes are located at

$$\cos \alpha_{xs} - \cos \alpha_x = \pm \frac{\lambda}{d_x} p$$

$$\cos \alpha_{ys} - \cos \alpha_y = \pm \frac{\lambda}{d_y} q$$

$$p, q = 0, 1, 2, \ldots$$

The lobe at $p = q = 0$ is the main beam. A triangular grid is more efficient for suppressing grating lobes than a rectangular grid,[44] so that for a given aperture size,

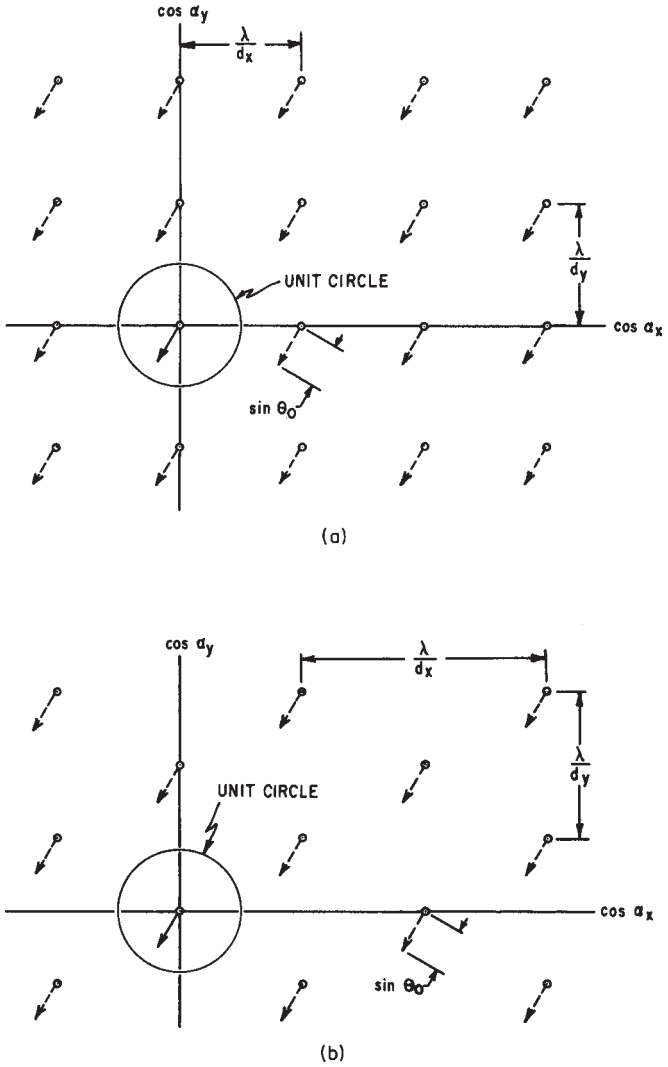


FIGURE 13.10 Grating-lobe positions for (*a*) rectangular and (*b*) triangular grids, showing the motion of the lobes as the beam is scanned at angle θ_0

fewer elements are required. If the triangular lattice contains elements at (md_x, nd_y), where $m + n$ is even, the grating lobes are located at

$$\cos \alpha_{xs} - \cos \alpha_x = \pm \frac{\lambda}{2d_x} p$$

$$\cos \alpha_{ys} - \cos \alpha_y = \pm \frac{\lambda}{2d_y} q$$

where $p + q$ is even.

Because only one main beam is normally desired in real space, an appropriate design will place all but one maximum in imaginary space for all angles of scan. With scanning, lobes that were originally in imaginary space may move into real space if the element spacing is greater than $\lambda/2$. As the array is scanned away from broadside, each grating lobe (in sin θ space) will move a distance equal to the sine of the angle of scan and in a direction determined by the plane of scan. To ensure that no grating lobes enter real space, the element spacing must be chosen so that for the maximum scan angle θ_m, the movement of a grating lobe by sin θ_m does not bring the grating lobe into real space. If a scan angle of 60° from broadside is required for every plane of scan, no grating lobes may exist within a circle of radius $1 + \sin \theta_m = 1.866$. The square grid that meets this requirement has

$$\frac{\lambda}{d_x} = \frac{\lambda}{d_y} = 1.866 \quad \text{or} \quad d_x = d_y = 0.536\lambda$$

Here, the area per element is

$$d_x d_y = (0.536\lambda)^2 = 0.287\lambda^2$$

For an equilateral-triangular array, the requirement is satisfied by

$$\frac{\lambda}{d_y} = \frac{\lambda}{\sqrt{3}\, d_x} = 1.866 \quad \text{or} \quad d_y = 0.536\lambda \quad d_x = 0.309\lambda$$

Because elements are located only at every other value of *mn*, the area per element is

$$2d_x d_y = 2(0.536\lambda)(0.309\lambda) = 0.332\lambda^2$$

For the same amount of grating-lobe suppression, the square geometry requires approximately 16% more elements.

Element-Phasing Calculations. A computer is usually required to perform the steering computations for a phased array antenna. It can compensate for many of the known phase errors caused by the microwave components, the operating environment, and the physical placement of the elements. For example, if the insertion and differential phase variations (which may occur from phase shifter to phase shifter) are known, they may be taken into account in the computations. Known temperature variations across the array that would induce phase errors may be compensated for. Finally, many feeds (e.g., optical and series feeds) do not provide equal phase excitation at the input to each phase shifter. The relative phase excitation caused by these feeds is a known function of frequency. In these cases, the computer must provide a correction based on the location of the element in the array and on the frequency of operation.

For a large array with thousands of elements, many calculations are required to determine the phasing of the elements. These calculations must be performed in a short period of time. The use of the orthogonal phase commands mT_{xs}, nT_{ys} helps to minimize these calculations. Once the element-to-element phase increments T_{xs}, T_{ys} have been computed for a given beam-pointing direction, the integral multiples of T_{ys} may be used to steer the columns (Figure 13.8).

13.4 APERTURE MATCHING AND MUTUAL COUPLING[45]

Significance of Aperture Matching. An antenna is a device that acts as a transformer to provide a good match between a source of power and free space. If the antenna is not matched to free space, power will be reflected back toward the generator, resulting in a loss in radiated power. In addition, a mismatch produces standing waves on the feed line to the antenna. The voltage at the peaks of these standing waves is $(1 + |\Gamma|)$ times greater than the voltage of a matched line, where Γ is the voltage reflection coefficient. This corresponds to an increased power level that is $(1 + |\Gamma|)^2$ times as great as the actual incident power. Therefore, while the antenna is radiating less power, individual components must be designed to handle more peak power. With antennas that do not scan, the mismatch may often be tuned out by conventional techniques, preferably at a point as close to the source of the mismatch as possible.

In a scanning array, the impedance of a radiating element varies as the array is scanned, and the matching problem is considerably more complicated. Unlike a conventional antenna, where mismatch affects only the level of the power radiated and not the shape of the pattern, spurious lobes in the scanning array may appear as a consequence of the mismatch. Further, there are conditions where an antenna that is well matched at broadside may have some angle of scan at which most of the power is reflected.

The variation in element impedance and element pattern is a manifestation of the mutual coupling between radiating elements that are in close proximity to one another. For a practical design, two empirical techniques are of great value:

1. Waveguide simulators provide a means for determining the element impedance in an infinite array with the use of only a few elements. The effectiveness of a matching structure based on these measurements may also be determined in the simulator.
2. A small array is the best technique for determining the active element pattern. The active element pattern, obtained by exciting one element and terminating its neighbors, is the best overall measure of array performance other than the full array itself. If a large reflection occurs at some angle of scan, it can be recognized by a null in the element pattern. The small array can also provide data on the coupling between elements. This data can be used to calculate the variation in impedance as the array is scanned.

Both these techniques will be discussed later in this section.

Effects of Mutual Coupling. When two antennas (or elements) are widely separated, the energy coupled between them is small and the influence of one antenna on the current excitation and pattern of the other antenna is negligible. As the antennas are brought closer together, the coupling between them increases. In general, the magnitude of the coupling is influenced by the distance between the elements, the pattern of the elements, and the structure in the vicinity of the elements. For example, the radiation pattern of a dipole has a null in the $\theta = \pm 90°$ direction and is omnidirectional in the $\theta = 0°$ plane. Therefore, it can be expected that dipoles in line will be loosely coupled and parallel dipoles will be tightly coupled. When an element is placed in an array of many elements, the effects of coupling are sufficiently strong that the pattern and impedance of the elements in the array are drastically altered.

The terms *active element pattern* and *element impedance* refer to an element in its operating environment (i.e., in an array with its neighboring elements excited). In the array, each excited element couples to every other element. The coupling from several elements to a typical central element, element 00, is shown in Figure 13.11. The $C_{mn,pq}$ are mutual-coupling coefficients relating the voltage (amplitude and phase) induced in the *mn*th element to the voltage excitation at the *pq*th element. The coupled signals add vectorially to produce a wave traveling toward the generator of element 00 that appears to be a reflection from the radiator of element 00. As the phases of the neighboring elements are varied to scan the beam, the vector sum of the coupled signals changes and causes an apparent change in the impedance of element 00. For some scan angles, the coupled voltages tend to add in phase, causing a large reflection and possibly the loss of the main beam. Large reflections often occur at scan angles just prior to the emergence of a grating lobe into real space, but in some instances such reflections may occur at smaller scan angles.

The description of the impedance variation given above made no reference to the feed network or the phase shifters and assumed that the only coupling between elements is via the radiating aperture. The coupling coefficients would be measured, and by superposition, the phased-voltage contributions from every element in the array (or at least those in the immediate vicinity) would be added vectorially to produce the voltage reflected back toward the generator. In a practical array, the impedance variation depends upon the feed system and the phase shifter. If these are taken into account, the impedance variation may be different from what the above model might predict. In most analyses, only the coupling at the aperture is considered. This description provides insight into the intrinsic impedance variation of the aperture when it is isolated from other effects, as in the case where each element has an independent feed (e.g., its own generator and isolator). In this case it, is a simple matter to measure the voltage-standing-wave ratio (VSWR) in any line and determine exactly the extent of the impedance and mismatch variation. For many feed systems, this is not possible, and a measurement of the reflected energy will provide erroneous information and a

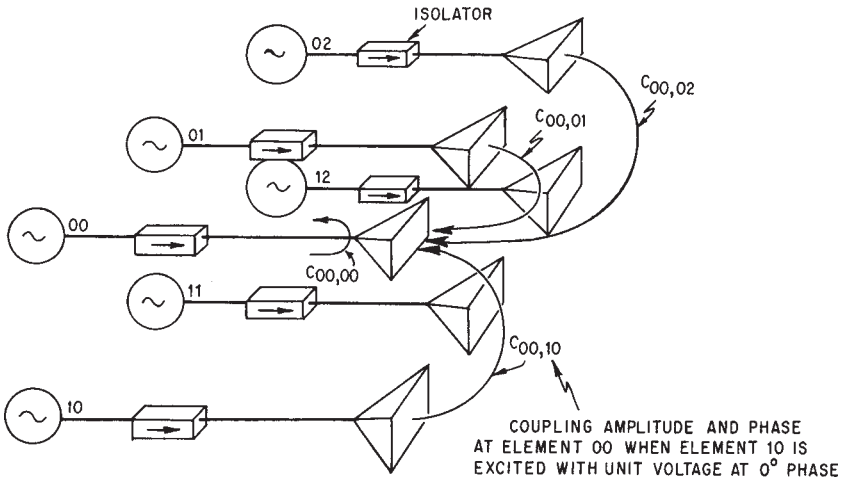


FIGURE 13.11 Coupled signals to a central element from neighboring elements

false sense of security. Unless all the reflections are collimated back at some central point (or independent feeds are used), some of the reflected energy will generally be re-reflected and contribute to undesirable sidelobes.

For large arrays, the impedance of an element located near the center of the array is often taken as typical of the impedance of every element in the array. As might be expected, this element is most strongly influenced by elements in its immediate vicinity. When the array is scanned, the influence of elements several wavelengths distant is also significant. For dipoles above a ground plane, the magnitude of the coupling between elements decays rapidly with distance. For a reasonable indication of array performance, an element in the center of a 5-by-5 array may be taken as typical of an element in a large array. For dipoles with no ground plane the coupling between elements does not decay so rapidly, and a 9-by-9 array appears reasonable. For an array of open-ended waveguides, a 7-by-7 array should suffice. If accurate prediction of the array performance is required, many more elements are needed than are indicated above.[47,48]

It is often convenient to assume that the array is infinite in extent and has a uniform amplitude distribution and a linear-phase taper from element to element. In this manner, every element in the array sees exactly the same environment, and the calculations for any element apply equally to all. These assumptions provide a significant simplification in the calculation of the element impedance variations. In addition, impedance measurements made in simulators correspond to the element impedance in an infinite array. In spite of the assumptions, the infinite-array model has predicted with good accuracy the array impedance and the impedance variations. Even arrays of modest proportions (less than 100 elements) have been in reasonable agreement with the results predicted for an infinite array.[49]

Element Pattern. From energy considerations, the directional gain of a perfectly matched array with constant amplitude distribution ($\eta = 1$) will vary as the projected aperture area from Eq. 13.14:

$$G(\theta_0) = \frac{4\pi A}{\lambda^2} \cos \theta_0$$

If it is assumed that each of the N elements in the array shares the gain equally, the gain of a single element is from Eq. 13.16:

$$G_e(\theta) = \frac{4\pi A}{N\lambda^2} \cos \theta_0$$

If the element is mismatched, having a reflection coefficient $\Gamma(\theta, \phi)$ that varies as a function of scan angle, the element gain pattern is reduced to

$$G_e(\theta) = \frac{4\pi A}{N\lambda^2} (\cos \theta)[1 - |\Gamma(\theta, \phi)|^2]$$

The element pattern is seen to contain information pertaining to the element impedance.[50–53] The difference between the total power radiated in the element pattern and the power delivered to the antenna terminals must equal the reflected power. In terms of the radiation pattern of the scanning array, this means that since the scanned antenna patterns trace out the element pattern, it follows that the average power lost from the scanned pattern is equal to the power lost from the element

pattern because of reflections. It is not enough to match one element in the presence of all its terminated neighbors. The element will deliver power to its neighbors, and this loss in power corresponds to the average power lost when scanning. An ideal, although not necessarily realizable element pattern, would place all the radiated power into the scan region, giving a pattern like a cosine on a pedestal and thereby providing maximum antenna gain for the number of elements used.

Thinned Arrays. The number of radiating elements in an array may be reduced to a fraction of those needed completely to fill the aperture without suffering serious degradation in the shape of the main beam. However, average sidelobes are degraded in proportion to the number of elements removed. The element density may be thinned so as to taper the amplitude distribution effectively, and the spacing is such that no coherent addition can occur to form grating lobes. A thinned aperture, where elements have been removed randomly from a regular grid,[54] is shown in Figure 13.12. The gain is that due to the actual number of elements $NG_e(\theta)$, but the beamwidth is that

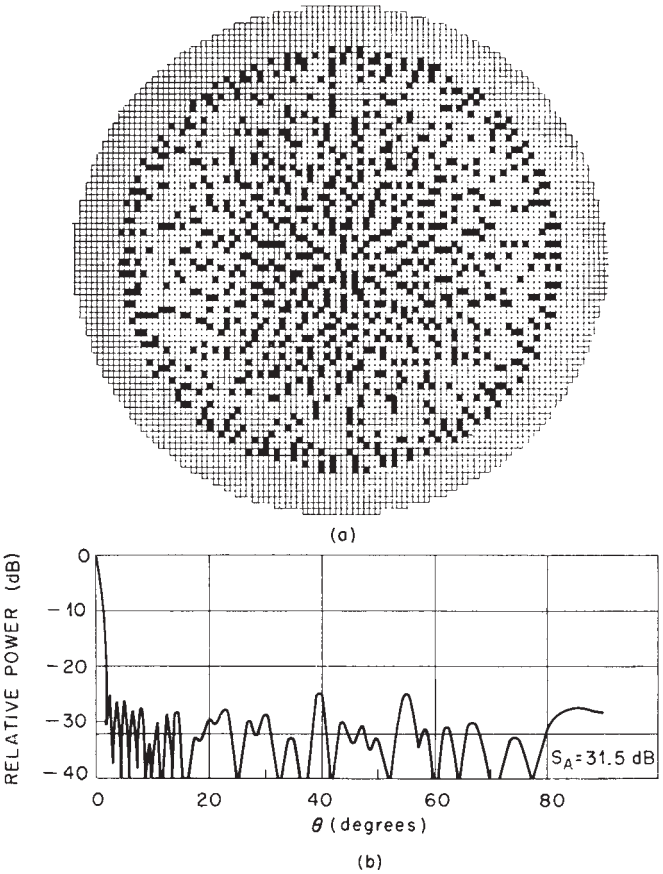


FIGURE 13.12 (*a*) Thinned array with a 4000-element grid containing 900 elements. (*b*) Pattern for a thinned array. S_A is the average sidelobe level. (*after R. E. Willey*[54] *© IRE 1962. Courtesy of Bendix Radio.*)

of the full aperture. For example, if the array has been thinned so that only 10% of the elements are used, the gain of the array will drop by 10 dB. However, because the main beam is virtually unchanged, about 90% of the power is delivered to the sidelobe region. Thinned arrays are seldom used.

If the removed elements (in a regular thinned array) are replaced with elements with matched loads, the element pattern is identical to that of one in the regular array with all elements excited. The element pattern is independent of the array excitation, and the same fractional amount of power will be lost (because of mismatch) whether the array is thinned, tapered, or uniformly illuminated. It should be noted that the concept of an element pattern that applies equally to every element is valid only when isolating feeds are used and edge effects are ignored.

A thinned array may also be implemented with an irregular element spacing, although this is not common. In this case, the element gain (and impedance) will vary from element, depending upon the environment of a given element. To obtain the gain of the array, it is necessary to sum all the different element gains $G_{en}(\theta)$. Thus

$$G(\theta) = \sum_n G_{en}(\theta)$$

Impedance Variation of Free Space. It is of interest to examine the case of a large continuous aperture that may be considered to be the limiting case of an array of many very small elements.[55] The free-space impedance E/H varies as cos θ for scanning in the E plane and as sec θ for scanning in the H plane. The impedance of a medium is thus dependent upon the direction of propagation, and the impedance variation of a scanning aperture is a natural consequence of this dependence. The continuous aperture appears to represent a lower limit to the impedance variation with scanning. This is indicated by Allen's results[56] where impedance variation with scanning was calculated for dipoles above a ground plane. In spite of increased mutual coupling, or perhaps because of it, the more closely the dipoles were spaced, the smaller the impedance variation with scanning. Although the impedance variation decreased, the absolute impedance of the dipoles also decreased, making them more difficult to match at broadside. It is expected that to obtain an impedance variation smaller than that of free space some impedance compensation must be employed.

Mutual Coupling and Surface Waves. The mutual coupling between two small isolated dipoles[57] should decrease as $1/r$ in the H plane and $1/r^2$ in the E plane (E and H planes are interchanged for slots). Coupling measurements[58] have shown that in the array environment the rate of decay is slightly greater than predicted above, indicating that some of the energy is delivered to other elements in the array and may be dissipated and reradiated from these elements. The same measurements have shown that the phase difference of the energy coupled to elements is directly proportional to their distance from the excited elements, indicative of a surface wave traveling along the array, leaking energy to each of the elements. For best performance, the velocity of the surface wave should be very close to that of free space. If the array contains waveguides or horns loaded with dielectric, the velocity will decrease slightly. Further, if the dielectric protrudes from the radiators or if a dielectric sheet is used in front of the array, the velocity of the surface wave may decrease dramatically. This surface wave is important because it can cause a large reflection (and an accompanying loss of the beam) for some angles of scan. This can best be seen by examining the condition of

phasing for which the couplings from many elements will add in-phase to cause a large reflection in a typical element.

Consider an array in which the velocity of the surface wave is that of free space. The difference in the phase of the voltages coupled from an adjacent pair of elements to element 00 (e_{00} in Figure 13.13) is related to the scan angle by

$$\frac{2\pi s}{\lambda} + \frac{2\pi s}{\lambda}\sin\theta_0 = \frac{2\pi s}{\lambda}(1+\sin\theta_0)$$

The couplings will be in phase when $\Delta\psi = 2\pi$ or when

$$\frac{s}{\lambda} = \frac{1}{1+\sin\theta_0}$$

This is seen to be exactly the same conditions as previously determined for the emergence of a grating lobe into real space. Therefore, it may be expected that when a grating lobe is about to emerge into real space, the coupled voltages tend to add in phase and cause a large mismatch.

Array Simulators. A good deal of effort has gone into matching a radiator in the presence of an array of radiators. The use of waveguide simulators as developed by Wheeler Laboratories has made it possible to determine the matching structure experimentally without needing to build an array. A waveguide, operating in a TE_{10} mode, may be considered to contain two inclined–plane waves propagating down the guide. The angle that each of the plane waves makes with the longitudinal direction (Figure 13.14) is determined by the H dimension of the waveguide and simulates the angle of scan of an infinite array

$$\sin\theta = \frac{\lambda}{\lambda_c} \tag{13.21}$$

where θ = scan angle
λ = free-space wavelength
λ_c = cutoff wavelength of guide

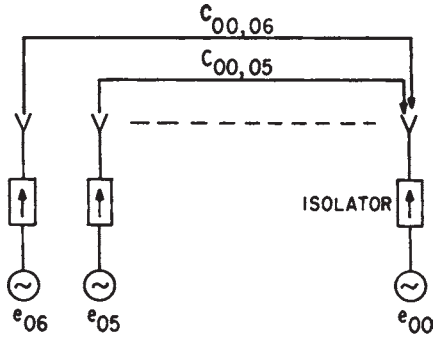


FIGURE 13.13 Two adjacent elements coupling to another element in the same row

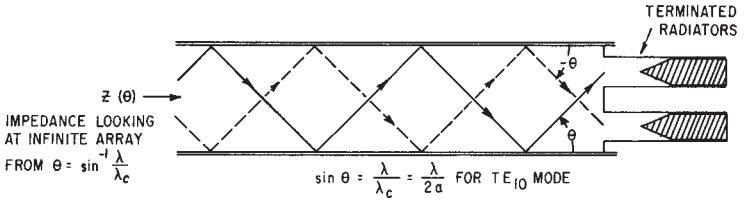


FIGURE 13.14 Array simulator terminated with two dummy elements

Additional scan angles may be simulated by exciting other modes. The waveguide dimensions are chosen so that a radiating element or element placed in the waveguide sees mirror images in the walls of the waveguide that appear to be at the same spacing as the array to be simulated. Both rectangular and triangular arrays may be simulated, as shown in Figure 13.15. The impedance measurements are made by looking into a waveguide simulator that is terminated with dummy elements. This is equivalent to looking at an infinite array from free space at a scan angle given by Eq. 13.21. A matching structure, designed from the simulator impedance data, may be placed into the simulator to measure its effectiveness. Several simulator designs, results, and a complete discussion of the topic have been presented by Hannan and Balfour.[59] The technique is limited in that only discrete scan angles can be simulated. Several scan angles in both planes of scan give a general idea of the array impedance.

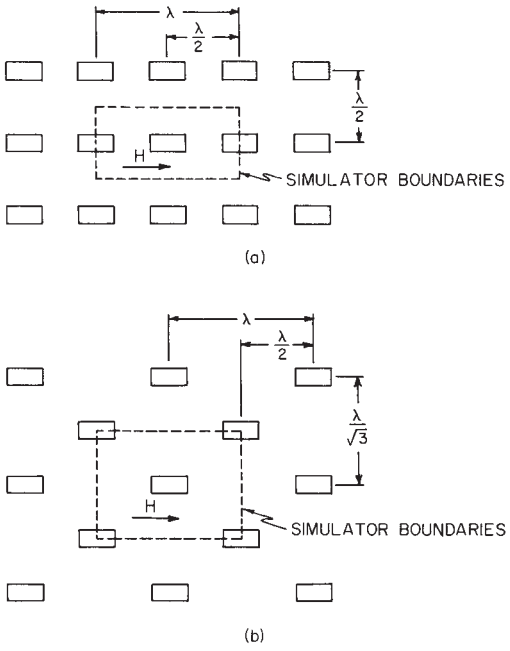


FIGURE 13.15 Rectangular- and triangular-array geometries with simulator boundaries superimposed: (*a*) square array with simulator superimposed and (*b*) triangular array with simulator superimposed

Compensation for Scanned Impedance Variation. The impedance of an element in an array has been discussed and has been shown to vary as the array is scanned. An array that is matched at broadside can be expected to have at least a 2:1 VSWR at a 60° angle of scan. To compensate for the impedance variation, it is necessary to have a compensation network that is also dependent on scanning.

Small Arrays. The element pattern is the best single indicator of impedance matching in a scanning array. One way of determining the element pattern is to build a small array. A central element is excited, and all other elements are terminated. The pattern of this central element is the active element pattern. Diamond[60] has examined the number of elements required in a small array to provide a reasonable approximation to an element in an infinite array. He concludes that 25 to 37 elements are required to provide a good indication. Figure 13.16 shows the change in the measured active element pattern as the number of elements is increased. For a 41-element array, the null is very pronounced. Even for the 23-element array, it is clear that the gain variation with scanning is dramatically greater than cos θ.

The small array can also be used to measure coupling coefficients as demonstrated by Figure 13.11 These coupling coefficients can be used to calculate the impedance variation as the array is scanned. Grove, Martin, and Pepe[61] have noted that for the element to be matched in its operating environment the self-coupling must exactly cancel the coupling from all other elements. They have used this technique to provide a good match on an ultralow-sidelobe wideband phased array.

The combination of waveguide simulators and small arrays provides powerful empirical tools to supplement the analytical techniques. Experience has demonstrated that a large antenna should not be built until the element pattern has been verified with a small array.

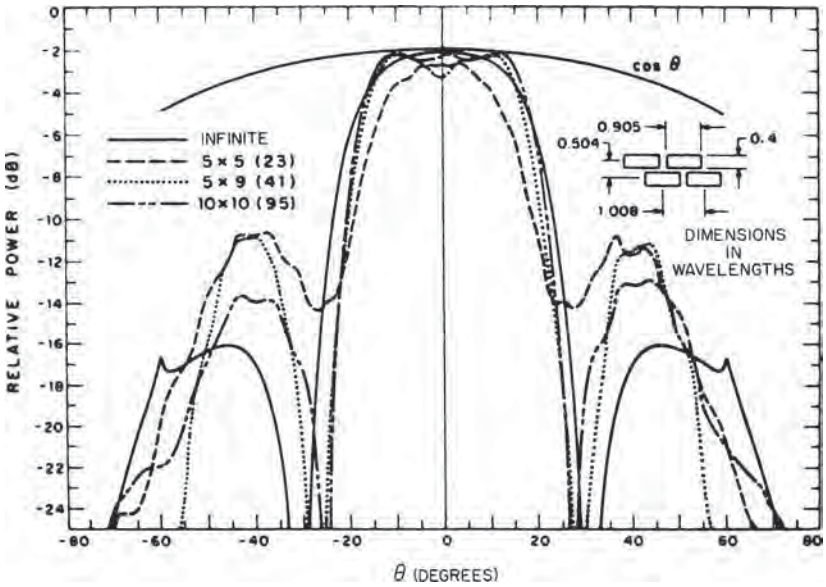


FIGURE 13.16 Experimental H-plane patterns of the center elements of waveguides arrays (*after B. L. Diamond*[60] *© Artech House 1972*)

13.5 LOW-SIDELOBE PHASED ARRAYS

Low sidelobes have long been of interest to antenna designers. This interest has been heightened by the jamming that threatens most military radars. The requirement for low sidelobes for clutter rejection in the AWACS radar resulted in technology that now supports sidelobe levels of more than 50 dB below the main-beam peak.[62,63] The price that must be paid to achieve these low sidelobes includes (1) a reduction in gain, (2) an increase in beamwidth, (3) increased tolerance control, (4) increased cost, and (5) the need to operate in an environment free from obstructions that can readily increase the sidelobes.[64] In spite of these drawbacks, the trend to low-sidelobe antennas has accelerated since low sidelobes provide a counter to electronic countermeasures (ECM).

Antenna sidelobes can be controlled by the aperture amplitude distribution. For phased arrays, the amplitude of each element may be controlled individually, and therefore, good sidelobe control can be achieved. The process of designing a low-sidelobe antenna can be considered in two parts:

1. Choose the correct illumination function to achieve the desired design (error-free) sidelobes.
2. Control the phase and amplitude errors that contribute to the random sidelobes.

Of the two, controlling errors fundamentally limits sidelobe performance. The effects of illumination function and errors are discussed below.

Illumination Functions. The relation between aperture illumination and the far-field pattern has been studied extensively and is well documented in the literature.[65–68] For a continuous aperture, the far-field pattern is the Fourier transform of the distribution across the aperture. For an array, samples are taken of the continuous distribution at each of the discrete locations. Some typical illumination functions are given in Table 13.1. It is seen that uniform illumination (constant amplitude) results in the highest gain and the narrowest beamwidth but at the cost of high sidelobes. As the amplitude is tapered, the gain drops, the beam broadens, and the sidelobes may be reduced. It is important for the antenna designer to choose an efficient and realizable illumination function that provides low sidelobes at a minimum loss in gain. For low-sidelobe radars, the Taylor illumination[69,70] for the sum patterns and the Bayliss illumination[71] for the difference patterns have almost become an industry standard. The Taylor illumination is somewhat similar to a cosine squared on a pedestal and is readily implemented. The Bayliss illumination is a derivative form of the Taylor illumination and is also readily implemented. It should be noted that in many phased arrays the sidelobe performance for the difference pattern is comparable to that of the sum pattern. For both sum and difference patterns, the sidelobes are referenced to the peak of the sum pattern. The beamwidth factor provides the beamwidth, in degrees, of an aperture with length a.

Figure 13.17 gives the approximate loss in gain and the beamwidth factor for the Taylor illumination as the sidelobes change. For a more comprehensive treatment, see Barton and Ward.[72] The sidelobes predicted by Table 13.1 are for antennas that have perfect phase and amplitude across the aperture. To allow for errors, aperture illuminations are often chosen to provide peak sidelobes below those required. For example, if the antenna specification calls for –40 dB sidelobes, a Taylor illumination that provides –45 dB design sidelobes might be chosen. The terminology $\bar{n}$ indicates that the first n sidelobes are held to the specified level.

TABLE 13.1 Illumination Functions

Illumination Function	Efficiency, η	Peak Sidelobe, dB	Beamwidth Factor, k
Linear illumination functions: beamwidth = $k\lambda/a$ (degrees); a = length of antenna			
Uniform	1	–13.3	50.8
Cosine	0.81	–23	68.2
Cosine square (Hanning)	0.67	–32	82.5
Cosine squared on 10 dB pedestal	0.88	–26	62
Cosine square on 20 dB pedestal	0.75	–40	73.5
Hamming	0.73	–43	74.2
Taylor $\bar{n} = 3$	0.9	–26	60.1
Taylor $\bar{n} = 5$	0.8	–36	67.5
Taylor $\bar{n} = 8$	0.73	–46	74.5
Circular illumination functions: beamwidth = $k\lambda/D$ (degrees); D = diameter of antenna			
Uniform	1	–17.6	58.2
Taylor $\bar{n} = 3$	0.91	–26.2	64.2
Taylor $\bar{n} = 5$	0.77	–36.6	70.7
Taylor $\bar{n} = 8$	0.65	–45	76.4

It should be noted that for a rectangular array a different illumination may be chosen for each plane. This is appropriate if the sidelobe requirements in each of the planes are different. The resultant loss in gain is then the sum (in decibels) of the losses in each plane.

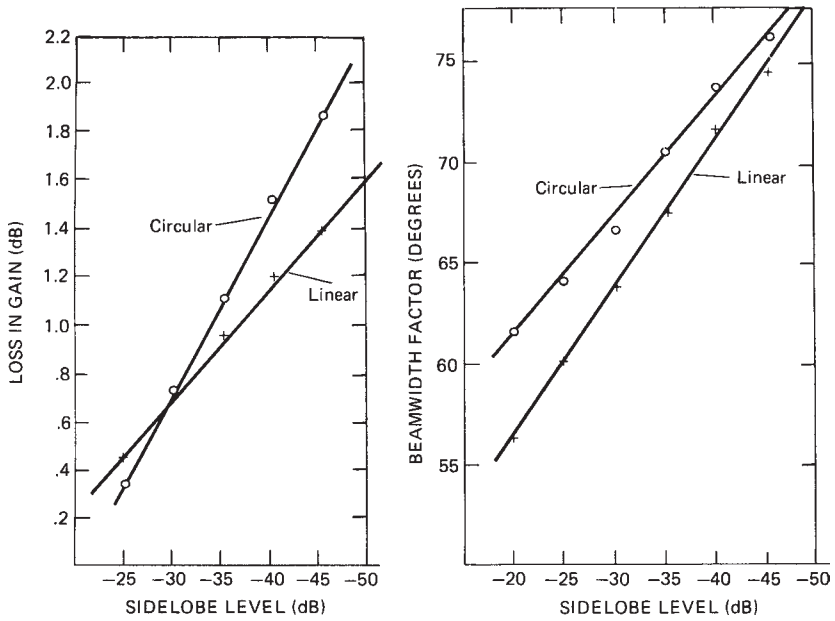


FIGURE 13.17 Taylor illumination: loss and beamwidth factor

Effect of Errors. When errors occur in phase or amplitude, energy is removed from the main beam and distributed to the sidelobes. If the errors are purely random, they create random sidelobes that are considered to be radiated with the gain and pattern of the element. When the errors are correlated, the sidelobe energy will be lumped at discrete locations in the far field. The correlated errors will, therefore, provide higher sidelobes, but only at a limited number of locations. Both correlated and random sidelobes are of concern to antenna designers. Correlated errors are discussed in Section 13.6.

Analyses of the far-field effect of errors are based on the fact that antennas are linear devices. That is, the far-field pattern is the sum of the voltage (amplitude and phase) of each radiating element in the antenna. For this reason, the far-field voltage pattern can be considered to be the sum of the design pattern plus the pattern created solely by the errors:

$$E_T(\theta,\phi) = E_{\text{design}}(\theta,\phi) + E_{\text{error}}(\theta,\phi)$$

In general, three regions will be recognized in the total resultant pattern: a low noise floor generated by random errors that follow the element pattern, a few peak sidelobes due to correlated errors, and the main beam with its sidelobes due to the design distribution.

Random Errors. Allen[73] and Ruze[74] have made detailed analyses of the effects of random errors on antennas. This discussion will follow the analysis performed by Allen. As previously mentioned, amplitude and phase errors take a fraction of the energy from the main beam and distribute this energy to the sidelobes. This fraction is, for small independent random errors,

$$\sigma_T^2 = \sigma_\phi^2 + \sigma_A^2$$

where σ_ϕ = rms phase error, rad
σ_A = rms amplitude error, volts/volt (V/V)

This energy is radiated into the far field with the gain of the element pattern. To determine the mean-squared-sidelobe level (MSSL), it is necessary to compare this energy with the peak of the pattern of an array of N elements so that the mean-squared-sidelobe level is

$$\text{MSSL} = \frac{\sigma_T^2}{\eta_a N(1-\sigma_T^2)} \tag{13.22}$$

Note that in the denominator of this expression the gain due to the array factor N is reduced by the aperture efficiency η_a and by the error power lost from the main beam $(1-\sigma_T^2)$. As an example, consider an array of 5000 elements with an aperture efficiency of 70%, $\sigma_a = 0.1$ v/v, and $\sigma_\phi = 0.1$ rad. Then $\sigma_T^2 = (0.1)^2 + (0.1)^2 = 0.02$.

$$\text{MSSL} = \frac{\sigma_T^2}{\eta N(1-\sigma_T^2)} = \frac{0.02}{(0.7)(5000)(0.98)} = 5.8 \times 10^{-6} = -52 \text{ dB}$$

The result is that this array has a mean floor of random sidelobes that on the average is 52 dB below the peak of the beam. It also illustrates that to achieve low sidelobes

very tight tolerances are required. The amplitude of 0.1 v/v is equivalent to a total amplitude standard deviation of 0.83 dB rms. The total rms phase error is 5.7°. It should be noted that there are numerous sources of phase and amplitude errors that are induced by the phase shifters, the feed network, the radiating elements, and the mechanical structure. The task of building a low-sidelobe antenna is one of reducing each of the amplitude errors to a few tenths of a decibel and the phase errors to a few degrees. The fewer the number of elements used, the tighter the tolerance becomes.

The individual effects of phase and amplitude errors and failed elements are summarized in Figure 13.18.[75] The resultant rms sidelobes are referenced to the gain of a single element so that the curve can be used for any number of elements with independent errors. For example, a 5° rms phase error will produce an rms sidelobe level that is approximately 21 dB below the gain of an element. If 1000 elements (30 dB) are used, the rms sidelobe level is 51 dB below the gain of the array. This is the effect of only the random phase errors. The effects due to amplitude errors and failed elements must also be included.

The previous discussion applies to the rms sidelobe level. This analysis has been extended by Allen[73] to apply the probability of keeping a single sidelobe below a given

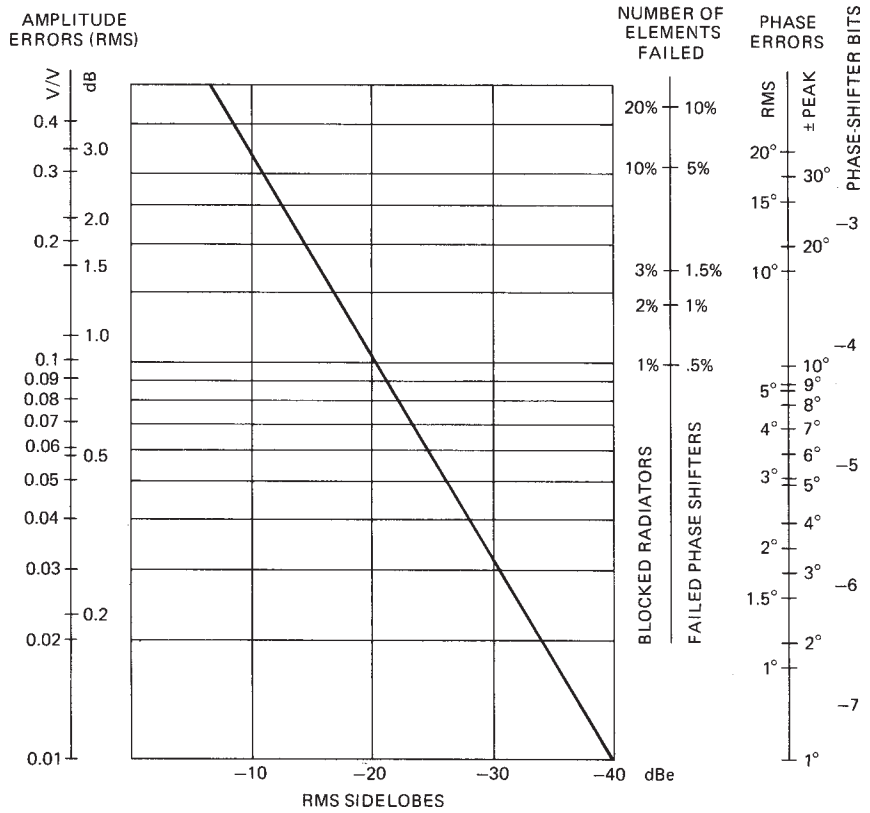


FIGURE 13.18 Random errors and rms sidelobes

level and then the probability of keeping a number of sidelobes below a given level. By ignoring the element pattern, the MSSL including failed elements is given by

$$\text{MSSL} = \frac{(1-P) + \sigma_A^2 + P\sigma_\phi^2}{\eta_a PN}$$

where $1 - P$ = probability of a failed element. Note that if $P = 1$ (no failed elements), this equation becomes

$$\text{MSSL} = \frac{\sigma_A^2 + \sigma_\phi^2}{\eta_a N} = \frac{\sigma_T^2}{\eta_a N}$$

This is the same as Eq. 13.22 except for $(1-\sigma_T^2)$ in the denominator, which is not significant for low-sidelobe antennas. For the case in which the design sidelobes are well below the sidelobes caused by errors, Allen has developed the set of curves shown in Figure 13.19. An example will illustrate the use of these curves. If you desire to hold the sidelobe at a given point in space to less than 40 dB below the peak of the beam with a probability of 0.99, draw a vertical line from –40 dB on the abscissa until it intersects the 0.99 curve. From this intersection, draw a horizontal line and read the value of MSSL, in this case –47 dB. Then

$$\text{MSSL} = -47 \text{ dB}$$

or

$$\text{MSSL} = 2 \times 10^{-5} = \frac{(1-P) + \sigma_A^2 + P\sigma_\phi^2}{\eta_a PN}$$

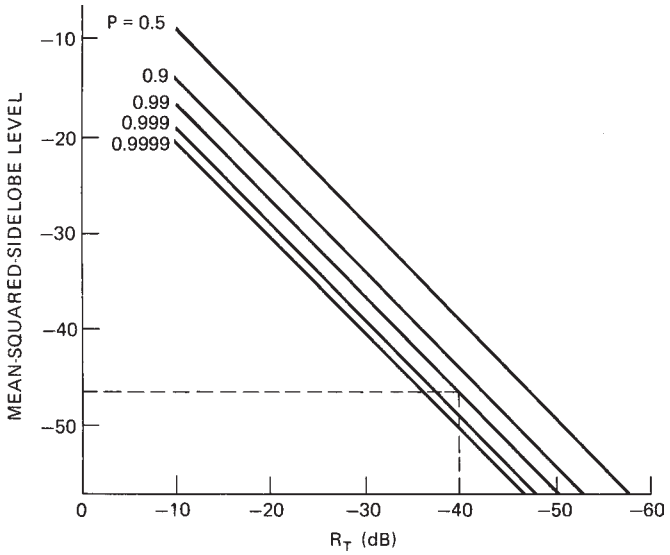


FIGURE 13.19 Sidelobe level to be held with probability P (*after J. L. Allen*[73])

For a 10,000-element array,

$$0.2 = \frac{(1-P) + \sigma_A^2 + P\sigma_\phi^2}{\eta_a P}$$

For $\eta_a = 1$, this array can tolerate $P = 0.83$, or $\sigma_A = 3.2$ dB, or $\sigma_\phi = 25.6°$. Naturally, each type of error must be anticipated, and one must allow a budget for failed elements, amplitude errors, and phase errors.

For a number of independent sidelobes, the probability that n sidelobes can be kept below a given level R_T is equal to the product of the probabilities that each sidelobe can be held below this level:

$$P\,[n \text{ sidelobes} < R_T] = \prod_{i=1}^{n} P[R(\theta_i) < R_T]$$

$$R(\theta_i) = \text{sidelobe level at } \theta_i$$

Assuming the same sidelobe requirement at each θ_i,

$$P\,[n \text{ sidelobes} < R_T] = \{1 - P\,[\text{one sidelobe} > R_T]\}^n$$

and for $P\,[\text{one sidelobe} > R_T] \ll 1$,

$$P\,[n \text{ sidelobes} < R_T] \cong 1 - n\,P\,[\text{one sidelobe} > R_T]$$

A simple example will illustrate the process. If it is necessary to keep all 100 sidelobes in a sector below –40 dB with a probability of 0.9, determine the required probability on any one given sidelobe:

$$0.9 = P\,[100 \text{ sidelobes} < 40 \text{ dB}]$$

Then

$$0.9 = 1 - 100\,P\,[\text{one given sidelobe} > 40 \text{ dB}]$$

$$0.001 = P\,[\text{one given sidelobe} > 40 \text{ dB}]$$

$$0.999 = P\,[\text{one given sidelobe} < 40 \text{ dB}]$$

That is, to keep all 100 sidelobes below –40 dB with a probability of 0.9, it is necessary to keep any given sidelobe below –40 dB with a probability of 0.999. The process of controlling each and every sidelobe is thus seen as formidable, since the total number of sidelobes is approximately equal to the number of elements in a phased array. For a 5000-element array and a probability of 0.999 that a single sidelobe will not exceed R_T at any single location, it will still be expected that 5 sidelobes will exceed R_T when all 5000 sidelobe locations are taken into account.

For very low sidelobe arrays, it is reasonable to allow a few sidelobes to exceed the MSSL value by as much as 10 to 12 dB to account for random variations. This can be seen from Figure 13.19 as the difference between $P = 0.5$ and $P = 0.999$ or $P = 0.9999$. If this allowance is not granted, the antenna will be overdesigned. It is worthwhile to do some probability calculations before specifying the exact sidelobe requirements.

13.6 QUANTIZATION EFFECTS

Of concern here are errors peculiar to phased arrays, which are due to the quantization of amplitude and phase and to the lobes that occur when these errors are repeated periodically. The effect on the gain and radiation pattern of random errors in the antenna excitation function is discussed in Section 13.5.

Phase Quantization. Phase shifters suitable for steering phased arrays are described in Section 13.9. Most of these phase shifters are digitally controlled and can be set with an accuracy that is a function of the number of bits. A small number of bits is desired for simplicity of phase-shift computation and operation, for minimal insertion loss in the case of diode phase shifters, and for minimal cost. On the other hand, a large number of bits is required for best performance in terms of gain, sidelobes, and beam-pointing accuracy.

Phase Errors. The phase of a phase shifter having P bits can be set to the desired value with a residual error:

$$\text{Peak phase error} = \alpha = \pm\frac{\pi}{2^P} \tag{13.23}$$

$$\text{RMS phase error} = \sigma_\phi = \frac{\pi}{2^P\sqrt{3}} \tag{13.24}$$

Loss in Gain. As discussed in Section 13.5, the loss in gain is $\sigma_\phi{}^2$, which with Eq. 13.24 gives

$$\Delta G \approx \sigma_\phi{}^2 = \frac{1}{3}\frac{\pi^2}{2^{2P}} \tag{13.25}$$

With many array elements, this result is statistically independent of the amplitude distribution. An enumeration of Eq. 13.25 gives

Number of phase-shifter bits, P	2	3	4
Loss in gain ΔG, dB	1.0	0.23	0.06

From the point of view of gain, therefore, 3 or 4 bits would appear ample.

RMS Sidelobes. Phase quantization decreases the gain of the main beam, as shown above. The energy that has been lost is distributed to the sidelobes. The resulting rms sidelobes are, therefore, σ_ϕ^2 relative to the gain of a single element, as shown in Figure 13.18.

Beam-Pointing Accuracy. The accurate determination of the direction of targets is made with the monopulse difference pattern. The accuracy of the null position of the difference pattern is, therefore, of interest. With quantized phase shifters, the position of this null can be moved with a granularity that is a function of the bit size.

Following the analysis of Frank and Ruze,[76] Figure 13.20 shows an aperture with an even number of elements N, separated by distances s. All elements are excited with

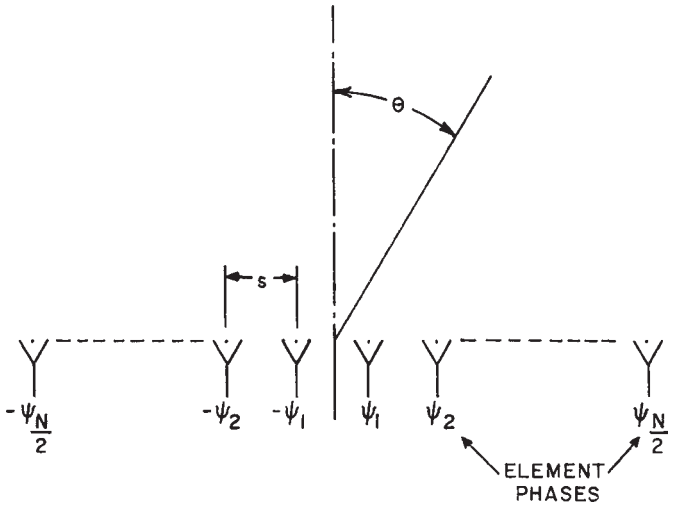


FIGURE 13.20 Antisymmetrically phased array

equal amplitude and antisymmetric phase to give a difference pattern. The steering increment is shown to be

$$\delta\theta = \frac{9}{N2^P}\theta_B(\text{scanned})$$

Periodic Errors

Periodic Amplitude and Phase Modulation. Both amplitude and phase quantization lead to discontinuities that may be periodic and give rise to *quantization lobes* that are similar to grating lobes.

Amplitude or phase errors that vary cosinusoidally may be analyzed simply after Brown.[77] Figure 13.21*a* shows an original amplitude distribution $F(x)$ disturbed by a cosinusoidal ripple $q \cos (2\pi x/s)$, giving a new distribution $F'(x)$ such that

$$F'(x) = F(x) + qF(x)\cos\frac{2\pi x}{s}$$

$$= F(x) + \frac{q}{2}\left[F(x)e^{j(2\pi x/s)} + F(x)e^{-j(2\pi x/s)}\right]$$

When the beam is scanned to θ_0, the quantization lobes occur at an angle θ_1, where

$$\sin\theta_1 = \sin\theta_0 \pm \frac{1}{s/\lambda}$$

The gain of the aperture varies as $\cos\theta$, and the relative amplitude of the quantization lobe is modified by the factor $\sqrt{(\cos\theta_1)/(\cos\theta_0)}$. Figure 13.21*b–g* shows the effects of various other periodic aperture modulations.

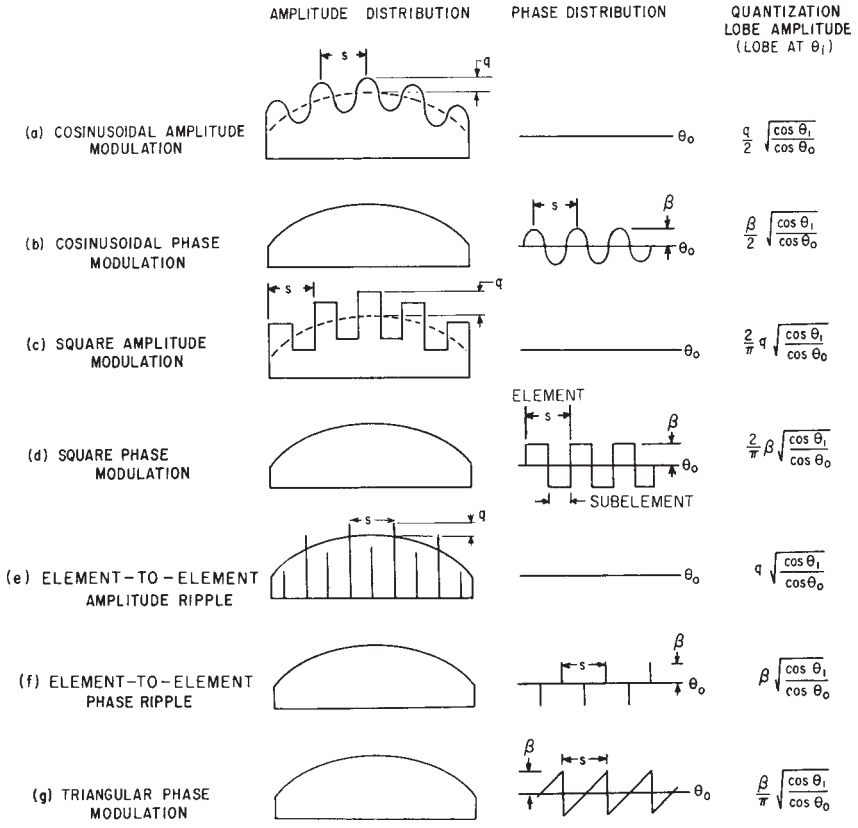


FIGURE 13.21 Effects of periodic amplitude and phase modulation ($\sin \theta_1 = \sin \theta_0 \pm \lambda/s$)

Phase-Shifter Quantization Lobes. Peak quantization sidelobe values are derived by Miller[78] considering the actual aperture phase distribution. Figure 13.22 shows this distribution for some scan angle θ_0 and the resulting errors due to phase quantization. Although a continuous curve has been drawn, only points corresponding to integral values of M are meaningful.

The greatest phase-quantization lobes seem to occur with a phase slope such that the elements are spaced by a distance exactly one-half of the period or an exact odd multiple thereof. The phase error under this condition has element-to-element phase ripples with a peak-to-peak value of $\alpha = \pi/2^p$. The value of the peak phase-quantization lobes for the case shown in Figures 13.21*f* and 13.22 are given in Figure 13.23. An examination of Figure 13.23 shows that peak phase-quantization lobes are significant and attempts should be made to reduce them.

Reduction of Peak Phase-Quantization Lobes. Miller[78] points out that peak quantization lobes can be reduced by decorrelating the phase-quantization errors. This may be done by adding a constant phase shift in the path to each radiator, with a value that differs from radiator to radiator by amounts that are unrelated to the bit size.

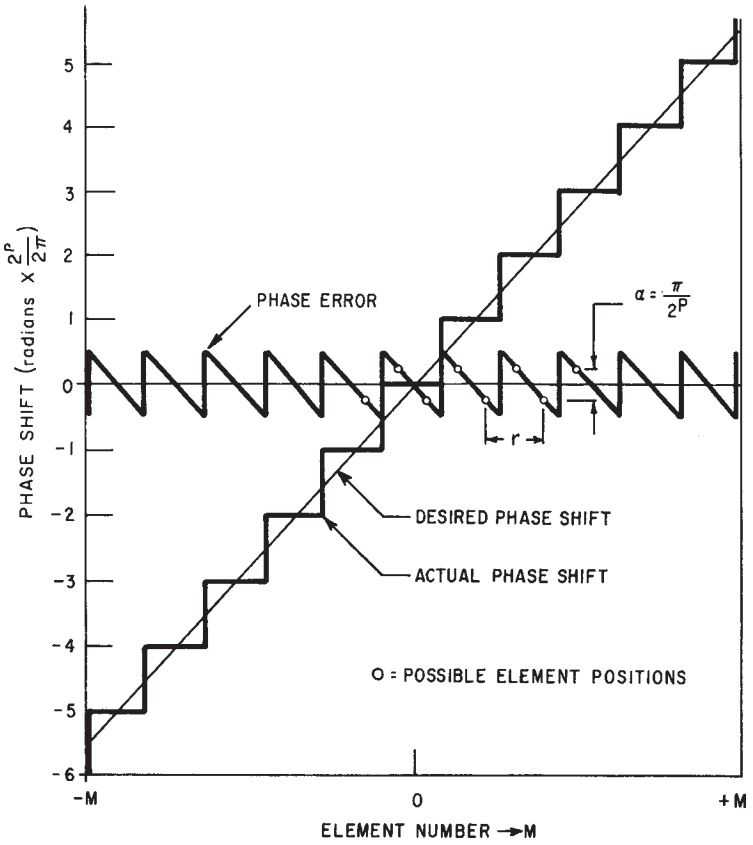


FIGURE 13.22 Aperture phase error due to phase quantization (*from Miller*[78])

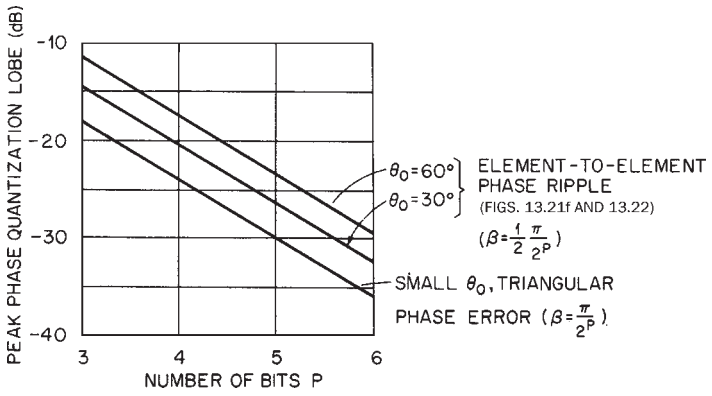


FIGURE 13.23 Peak sidelobes due to phase quantization

The variable phase shifter is then programmed to account for this additional insertion phase. With a spherical or quadratic law of insertion phase variation, as obtained with optical-feed systems (Section 13.8), the reductions in peak quantization lobes are equivalent to adding 1 bit to the phase shifters in a 100-element array, 2 bits in a 1000-element array, and 3 bits in a 5000-element array.

Amplitude Quantization. When the aperture of a phased array is divided into equal subarrays, then the amplitude distribution across each subarray is constant. Aperture taper for the antenna is approximated by changing the amplitude from subarray to subarray, and quantization lobes arise from these discontinuities. The value of these lobes may be estimated from the various results shown in Figure 13.21 or actually computed by summing all contributions at the known quantization (grating) lobe angles. The distribution becomes smoother as the number of subarrays is increased or as they are interlaced.

13.7 BANDWIDTH OF PHASED ARRAYS

The phenomenon of focusing an array is the result of the energy from each element adding in-phase at some desired point within the antenna. When energy is incident normal to the array, each element receives the same phase independent of frequency. When energy is incident from some angle other than normal, the phase difference from the planar phase front to each element is a function of frequency and most phased arrays with phase shifters become frequency-dependent. This same phenomenon can be viewed in the time domain. As shown in Figure 13.24, when a pulse of energy is incident at an angle other than normal, the energy is received earlier at one edge of the array than at the other edge and a period of time must elapse before energy appears in all elements. The concept of aperture fill time, $T = L/c \sin \theta_0$, is just another way of explaining the bandwidth of a phased array.

The bandwidth of phased arrays is described by Frank[79] as being composed of two effects: the aperture effect and the feed effect. In both effects, it is the path-length differences that contribute to the bandwidth sensitivity of a phased array. For a parallel-fed

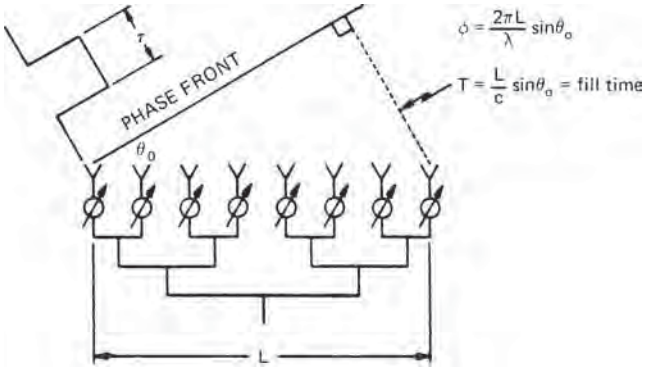


FIGURE 13.24 Aperture-fill-time parallel-fed array (*after J. Frank*[79] *© Artech House 1972*)

array (equal line length), the feed network does not contribute to a change in phase with frequency, and so only the aperture effect remains. This will be discussed first, and the effect of the feed will follow.

Aperture Effects. When energy is incident on an array at an angle other than broadside (Figure 13.24), the phase required on the edge element is $\psi = (2\pi L)/\lambda \sin \theta_0$. Note the λ in the denominator. This indicates that the required phase is frequency-dependent. If frequency is changed and the phase shifters are not changed, the beam will move. For an equal-line-length feed, the beam shape will be undistorted (in sine space) and the beam will move toward broadside as the frequency is increased. If the phase shifters are replaced by time-delay networks, then the phase through the time-delay networks will change with frequency and the beam will remain stationary.

When phase shift (independent of frequency) is used to steer the beam, the beam is steered to a direction θ_0 with a phase of

$$\psi = \frac{2\pi x}{\lambda_1} \sin \theta_0 = \frac{2\pi x}{c} f_1 \sin \theta_0$$

on the element located a distance x from the array center. At frequency f_2 this same phase setting steers the beam to a new direction, $\theta_0 + \Delta\theta_0$, which is shown in Frank[79] to be

$$\Delta\theta_0 = \frac{\Delta f}{f} \tan \theta_0 \qquad \text{(rad)} \tag{13.26}$$

As the frequency is increased, the beam scans toward broadside by an angle that is independent of the aperture size or beamwidth. However, the permissible amount that the beam may scan with frequency is related to the beamwidth because pattern and gain deteriorations are a function of the fractional beamwidth scanned. The angle that the beam actually scans, on the other hand, is related to the percentage bandwidth. A bandwidth factor may, therefore, be defined in terms of the broadside beamwidth:

$$\text{Bandwidth factor} = K = \frac{\text{bandwidth}\,(\%)}{\text{beamwidth}\,(^\circ)}$$

A reasonable criterion is to limit the bandwidth so that the beam never scans by more than ± one-fourth of a local beamwidth frequency, i.e.,

$$\text{Criterion:} \left| \frac{\Delta\theta_0}{\theta_B \text{ (scanned)}} \right| \le \frac{1}{4}$$

With a scan of 60°, this gives $K = 1$, and in terms of broadside beamwidth the limit is

$$\text{Bandwidth}\,(\%) = \text{beamwidth}\,(^\circ) \quad \text{(CW)}$$

For example, if the array has a beamwidth of 2° this criterion permits a 2% change in frequency prior to resetting the phase shifters. This allows the beam to move from one-fourth beamwidth on one side of the desired direction to one-fourth beamwidth on the other side as the frequency changes by 2%. At smaller scan angles, the effect is reduced as given by Eq. 13.26, and broader-band operation is possible.

The explanation given above applies to an antenna operating at a single (CW) frequency and describes how the beam moves as this frequency is changed.

However, most radars are pulsed and radiate over a band of frequencies. For a beam scanned from broadside, each spectral component is steered to a slightly different direction. To determine the composite effect of the components, it is necessary to add the far-field patterns of all spectral components. This analysis has been performed[80,81] and it is apparent that the overall antenna gain of the pulse will be less than that of a single spectral component that has maximum gain in the desired direction. As with the CW situation, the greatest loss occurs at the maximum scan angle, which is assumed to be 60°. For this situation, the criterion chosen is to allow a spectrum that loses 0.8 dB of energy on target due to frequency-scanned spectral components. For a beam scanned to 60°, this becomes

$$\text{Bandwidth (\%)} = 2 \text{ beamwidth (°)} \quad \text{(pulse)}$$

Note that this is twice the bandwidth permitted for the CW situation. Another way of analyzing this phenomenon is in terms of aperture fill time. As shown in Figure 13.24, the time it takes to fill the aperture with energy is

$$\tau = \frac{L}{c} \sin \theta_0$$

If the bandwidth is chosen to be equal to the aperture fill time $\tau = L/c \sin \theta_0$, this is equivalent to bandwidth (%) = 2 beamwidth (°). Hence, a loss of 0.8 dB can be expected if the pulse width is equal to the aperture fill time. Longer pulses will have less loss. The exact amount of the loss will depend on the specific spectrum transmitted, but the variation will amount to less than 0.2 dB for most waveforms. Rothenberg and Schwartzman[82] provide details and also treat the problem as a matched filter.

The preceding discussion assumes an equal-path-length feed. However, it is unlikely that a feed will provide exactly equal path lengths. It will suffice to have the path lengths kept within one wavelength of each another. The phase errors introduced can then be corrected by programming the phase shifters. This will have the beneficial effect of breaking up the quantization errors and thereby reducing quantization lobes.

Feed Effect. When an equal-path-length feed is not in use, the feed network will produce a change in phase with frequency. In some cases, such as Rotman lens[83] or equal-length Blass matrix[84], the feed can actually compensate for the aperture effect and produce a beam direction that is independent of frequency. However, the more conventional feeds tend to reduce the bandwidth of the array.

End-Fed Series Feed. An end-fed series array is shown in Figure 13.31*a* later in this chapter. The radiating elements are in series and progressively farther and farther removed from the feed point. When the frequency is changed, the phase at the radiating elements changes proportionately to the length of feed line so that the phase at the aperture tilts in a linear manner and the beam is scanned. This effect is useful for frequency-scanning techniques, but in the case of phased arrays, it is undesirable and reduces the bandwidth. It has been shown previously that with phased arrays, the pointing direction of a scanned beam also changes with frequency (Eq. 13.26) because phase rather than time delay is adjusted. These two changes in beam-pointing direction may add or subtract, depending on the direction in which the beam has been scanned. The worst case will be considered here.

With a change in frequency, a nondispersive transmission line having free-space propagation characteristics and a length L equal to the size of the aperture that it feeds will produce a linear-phase variation across the aperture with a maximum value at the edges of

$$\Delta\psi_{\max} = \frac{\Delta f}{f}\frac{2\pi L}{\lambda} \qquad \text{rad}$$

where $\Delta f/f$ is the fractional change in frequency. This linear-phase progression across the aperture will scan the beam by

$$\Delta\theta_0 = \frac{\Delta f}{f}\frac{1}{\cos\theta_0} \qquad \text{rad}$$

For one direction of scanning, this effect will add to the aperture effect; for the opposite direction, it will tend to cancel the aperture effect. In a waveguide, with the guide wavelength denoted λ_g, the effect is more pronounced and the resultant change in beam position is

$$\Delta\theta_0 = \frac{\lambda_g}{\lambda}\frac{\Delta f}{f}\frac{1}{\cos\theta_0} \qquad \text{rad}$$

When analyzing an end-fed series feed, it is necessary to consider both the feed effect and the aperture effect. The total frequency scan of this feed will be

$$\Delta\theta_0 = \frac{\Delta f}{f}\tan\theta_0 \pm \frac{\lambda_g}{\lambda}\frac{\Delta f}{f}\frac{1}{\cos\theta_0} \qquad \text{rad} \qquad \text{(CW)}$$

Center-Fed Series Feed.[85] A center-fed array (Figure 13.25) can be considered as two end feeds. Each feed controls an aperture that is half the total and, therefore, has twice the beamwidth. As the frequency is changed, each half of the aperture scans in the opposite direction. This initially creates a broader beam with reduced gain. As frequency continues to change, the two beams will eventually split apart. At broadside, the center-fed antenna has poorer performance than a parallel feed since each half scans. However, at 60° scan the compensation on one-half of the array assists in keeping the gain comparable to that of a parallel feed. From the viewpoint of gain reduction, the criterion for a center-fed array is

$$\text{Bandwidth}\,(\%) = \frac{\lambda}{\lambda_g}\,\text{beamwidth}\,(^\circ) \quad \text{(CW)}$$

where λ_g is the waveguide wavelength.

However, from the viewpoint of sidelobes, this criterion may be unacceptable. For a low-sidelobe design, either CW or pulsed, the sidelobes for this feed should be calculated because a change in frequency no longer produces a translated beam but rather a broader beam composed of two translated beams.

Space Feed. The space (optical) feed can be considered to be somewhat between a parallel feed and a center-fed series feed. With a very long focal length, the space feed approximates a parallel feed. With a very short focal length it approximates a

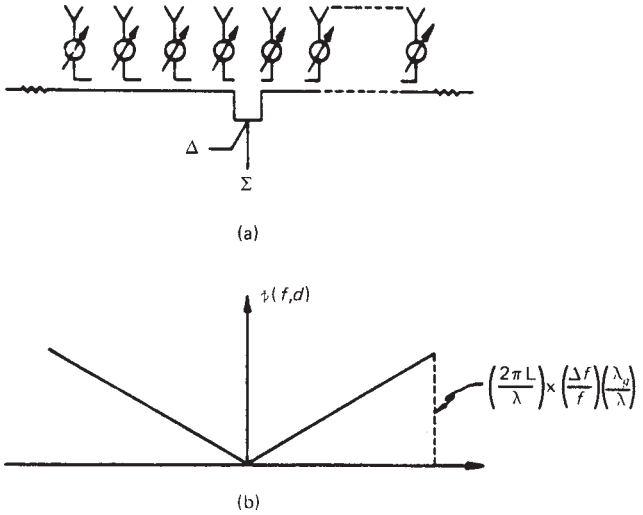


FIGURE 13.25 Center-fed series array (*after J. Frank*[79] © *Artech House 1972*): (*a*) center-fed series feed and (*b*) gabled phase due to feed

center-fed series feed. Because the bandwidth performance of each is comparable, the optical-feed criterion is

$$\text{Bandwidth (\%)} = \text{bandwidth (°)} \quad \text{(CW)}$$

Table 13.2 summarizes the bandwidth criteria for the various feed networks. Both the CW criterion and the comparable pulse criterion are given.

Broad Instantaneous Bandwidth. For a stationary beam in space, independent of frequency, it is necessary to use time delay rather than phase steering. It is not practical to provide a time-delay network at each element in a phased array because

TABLE 13.2 Bandwidth Criteria for Several Feed Networks*

Feed	CW Bandwidth (%)	Pulse Bandwidth (%)
Equal line length	Beamwidth	2 × beamwidth
End-fed series	$\frac{1}{\left(1+\frac{\lambda_g}{\lambda}\right)}\times$ beamwidth	$\frac{2}{\left(1+\frac{\lambda_g}{\lambda}\right)}\times$ beamwidth
Center-fed series	$\frac{\lambda}{\lambda_g}\times$ beamwidth	$2\,\frac{\lambda}{\lambda_g}\times$ beamwidth
Space-fed (optical)	Beamwidth	2 × beamwidth

**After Frank*[79]
NOTE: All beamwidths are in degrees and refer to the broadside beamwidth: λ_g = guide wavelength and λ = free-space wavelength.

these networks are expensive and lossy and contain errors. An alternative is to use a broadband beam-switching technique such as the equal-line-length Blass matrix[84] or a Rotman lens.[83] For 2D scanning these techniques become quite complex.

Another technique for improving the bandwidth by a considerable factor is to use an array of subarrays. The radiating elements of a phased array may be grouped into subarrays where time-delay elements are added. This is shown in Figure 13.26. The antenna may be regarded as an array of subarrays. The subarray pattern forms the element factor; it is steered by phase shifters in the desired direction, but it scans with frequency as indicated by Eq. 13.26. The array factor is scanned by adjusting the frequency-independent time-delay elements. All subarrays are steered in the same manner. The total radiation pattern is the product of the array factor and the element factor. A change in frequency gives rise to grating lobes rather than shifts of the beam position. This can be seen from Figure 13.27. The subarray pattern is shown at the design frequency f_0 and is seen to have a null at the position of the grating lobe. As frequency is changed by δf, the pattern is scanned. It is shown dashed in a position where it has been scanned by a little more than half of its beamwidth. This is clearly too much, for the product of array and element factors gives two beams of equal amplitude.

The loss in gain and the magnitude of the grating lobe are functions of the fractional subarray beamwidth that has been scanned as a result of the change in frequency. The results may be expressed in terms of the bandwidth factor K (referred to as the *subarray broadside beamwidth*).

Figure 13.28 shows these values as functions of K for a scan of 60°.

The value previously used for scanning 60° was $K = 1$ and this value appears acceptable here, where the relevant beamwidth is the broadside beamwidth of the subarray. Thus, if the aperture is split into N subarrays in one plane, with time-delay networks at each subarray level, the bandwidth is increased by a factor of N. This same bandwidth criterion leads to a reduction in gain of about 0.7 dB and a grating lobe of about −11 dB at the edges of the band with 60° of scan. Interlacing of subarrays can reduce the grating lobes.

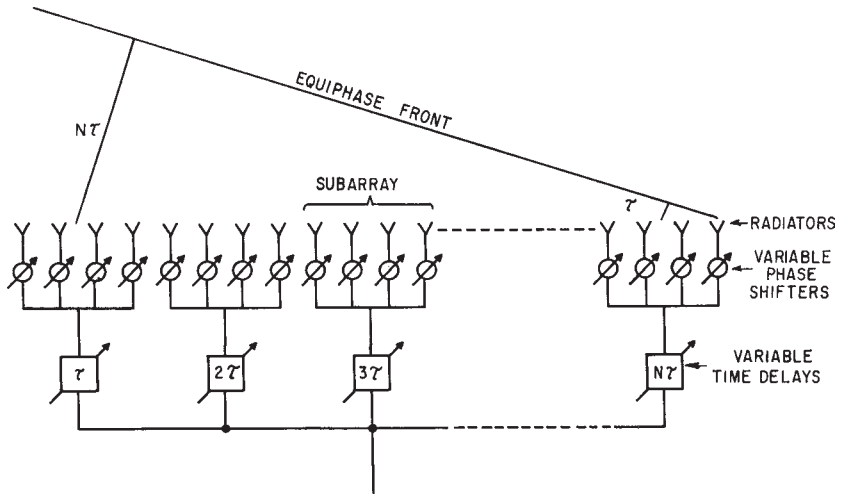


FIGURE 13.26 Phased array using subarrays with time delay

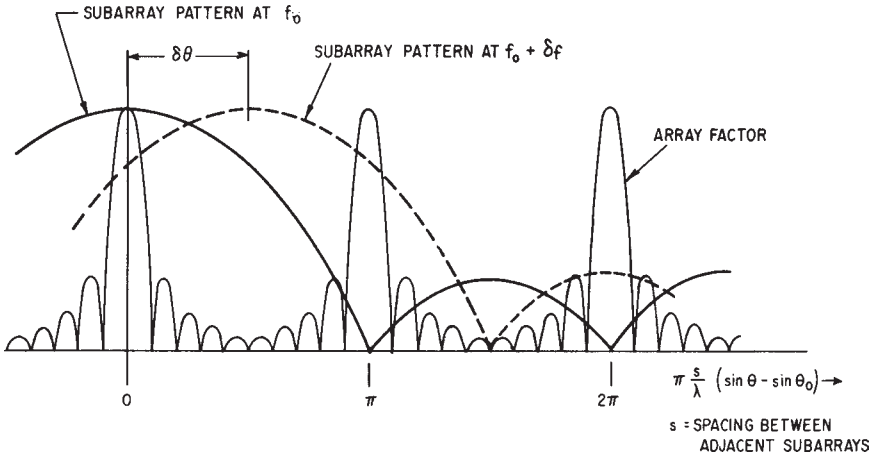


FIGURE 13.27 Generation of grating lobes by a change of frequency

The monopulse null position is unaffected by the behavior of the subarrays as long as they all respond in the same way. The null position is determined only by the time-delay networks behind the subarray and corresponds to the array-factor null, which is unaffected when multiplied by the subarray pattern.

Time-Delay Networks. Figure 13.29*a* shows a time-delay network that is digitally controlled by switches. The total delay path length that has to be provided non-dispersively amounts to $L \sin \theta_{max}$, where θ_{max} is the maximum scan angle for the aperture L. The smallest bit size is about $\lambda/2$ or λ, with the precise setting adjusted by

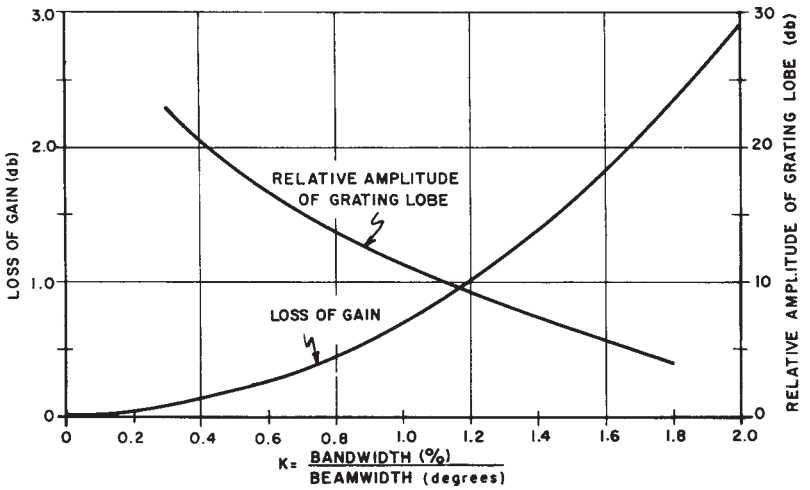


FIGURE 13.28 Loss of gain and grating-lobe amplitude as functions of bandwidth (phased subarrays with time delay, scanned 60°). The value of the grating lobe will be modified by the element pattern.

an additional variable phase shifter. A 1° beam scanned 60°, for example, requires a time delay of 6 bits, the largest being 32 wavelengths, as well as an additional phase shifter. The tolerances are tight, amounting in this case to a few degrees out of about 20,000 and are difficult to meet. Problems may be due to leakage past the switch, to a difference in insertion loss between the alternate paths, to small mismatches at the various junctions, to variations in temperature, or to the dispersive characteristics of some of the components. Painstaking design is necessary. The switches may be diodes or circulators. Leakage past the switches may be reduced by adding another switch in series in each line. The difference in insertion loss between the two paths may be equalized by padding the shorter arm. The various problems have been comprehensively assessed and analyzed by Lincoln Laboratory.[45]

On transmitting, the tolerances are less severe because the requirements are usually for power on target rather than for accurate angle determination or low sidelobes.

Figure 13.29*b* shows another configuration that has the advantage of simplicity. Each of the switchable circulators connects either directly across (counterclockwise) or via the short-circuited length. Isolation in excess of 30 dB is required. It is clear that the insertion loss of time-delay circuits is very high for most practical systems. They would, therefore, precede a final power amplifier for transmitting and follow a preamplifier for receiving.

A further method of providing delay is possible by translating the problem from the microwave domain and delaying at IF.

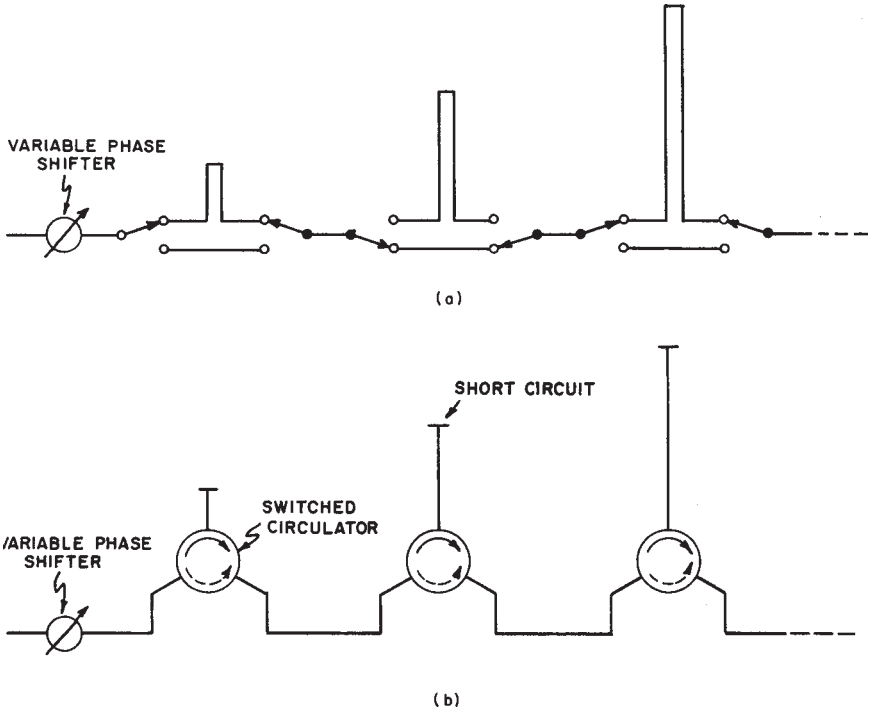


FIGURE 13.29 Time-delay configurations: (*a*) time delay by choosing upper or lower paths and (*b*) time delay by using switched circulators

13.8 FEED NETWORKS (BEAMFORMERS)

Optical-Feed Systems. Phased arrays may be in the form of lens arrays or reflectarrays, as shown in Figure 13.30, where an optical-feed system provides the proper aperture illumination. The lens has input and output radiators coupled by phase shifters. Both surfaces of the lens require matching. The primary feed should be designed with care and can be complex to give the desired aperture amplitude distribution with low spillover losses. The transmitter feed can be separated from the receiver feed by an angle α, as shown in Figure 13.30. The phase shifters are then reset between transmitting and receiving so that in both cases the beam points in the same direction. This method allows flexibility in optimizing the transmitter aperture distribution, perhaps for maximum power on the target, and separately optimizing the receiver sum and difference patterns for low sidelobes. Since a change in feed position corresponds to scanning with time delay, additional feeds may be added to provide several time-delay-compensated directions of scan for a corresponding increase in bandwidth.

The phasing of the antenna has to include a correction for the spherical phase front. This can be seen (Figure 13.30*a*) to amount to

$$\frac{2\pi}{\lambda}\left(\sqrt{f^2+r^2}-f\right)=\frac{\pi}{\lambda}\frac{r^2}{f}\left[1-\frac{1}{4}\left(\frac{r}{f}\right)^2+\ \ldots\right]$$

With a sufficiently large focal length, the spherical phase front may be approximated by that of two crossed cylinders, permitting the correction to be applied simply with row-and-column steering commands.

Correction of a spherical phase front with the phase shifters reduces peak phase-quantization lobes (Section 13.6). Space problems may be encountered in assembling an actual system, especially at higher frequencies, because all control circuits have to be brought out at the side of the aperture of the lens array.

Multiple beams may be generated by adding further primary feeds. All the beams will be scanned simultaneously by equal amounts in sin θ.

The reflectarray shown in Figure 13.30*b* has general characteristics similar to those of the lens. However, the same radiating element collects and radiates after reflection. Ample space for phase-shifter control circuits exists behind the reflector. To avoid aperture blocking, the primary feed may be offset as shown. As before, transmitting and receiving feeds may be separated. Multiple beams are again possible with additional feeds.

The phase shifter must be reciprocal so that there is a net controllable phase shift after passing through the device in both directions. This rules out frequently used nonreciprocal ferrite or garnet phase shifters.

Constrained Feeds. The optical-feed system divides power very simply in one step from the feed to the many elements on the aperture. In contrast, the constrained-feed system uses many steps. For a high-performance low-sidelobe system, each of these power-dividing steps has to be well matched over the band. If mismatches are present, they will be separated by many wavelengths and will add up to give frequency-sensitive phase and amplitude perturbations at the aperture. In general, these perturbations will be different for the monopulse sum and difference patterns, so that only a common average can be compensated by calibration.

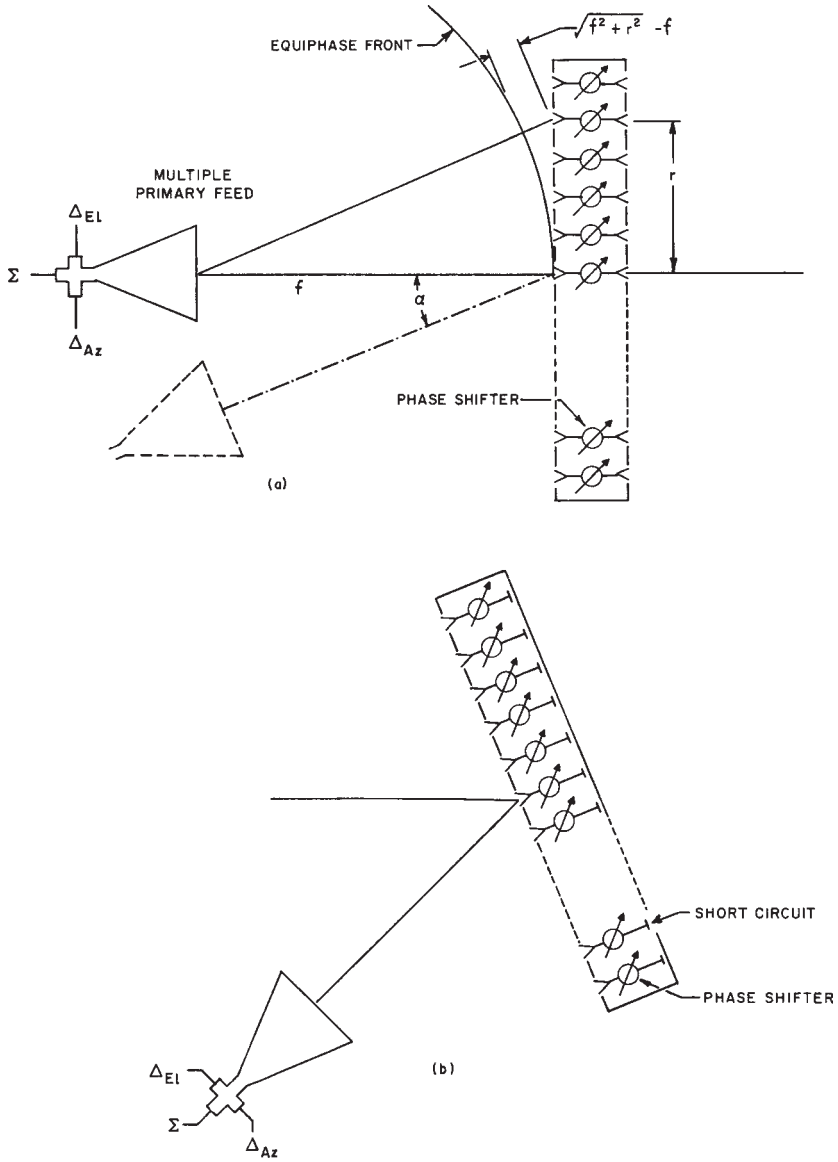


FIGURE 13.30 Optical-feed systems: (*a*) lens array and (*b*) reflectarray

Series Feeds. Figure 13.31 shows several types of series feeds. In all cases except *d*, the electrical path length to each radiating element has to be computed as a function of frequency and taken into account when setting the phase shifters. Type *a* is an end-fed array. It is frequency-sensitive and leads to more severe bandwidth restrictions than most other feeds. Type *b* is center-fed and has almost the same bandwidths as a

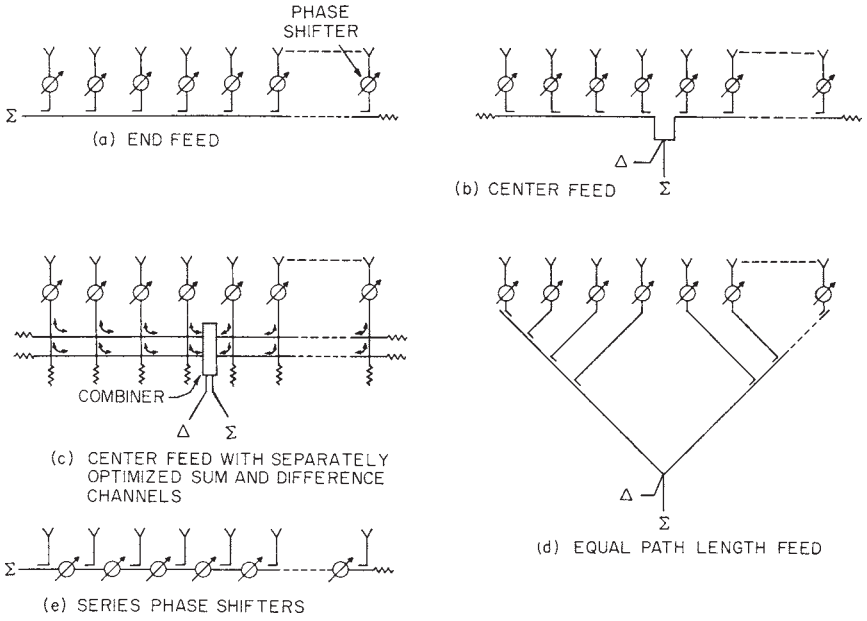


FIGURE 13.31 Series-feed networks

parallel-feed network. Sum- and difference-pattern outputs are available, but they have contradictory requirements for optimum amplitude distributions that cannot both be satisfied. As a result, either good sum or good difference patterns can be obtained. At the cost of some additional complexity, the difficultly can be overcome by the method shown in Figure 13.31*c*. Two separate center-fed feed lines are used and combined in a network to give sum- and difference-pattern outputs.[86] Independent control of the two amplitude distributions is possible. For efficient operation, the two feed lines require distributions that are *orthogonal*, that is, that give rise to patterns where the peak value of one coincides with a null from the other and aperture distributions are respectively even and odd.

A very wideband series feed with equal path lengths is shown in Figure 13.31*d*. If the bandwidth is already restricted by phase scanning, then very little advantage is obtained at the cost of a considerable increase in size and weight. The network in Figure 13.31*e* gives simplicity in programming because each phase shifter requires the same setting. The insertion loss increases for successive radiators, and the tolerances required for setting the phases are high. This type is not commonly used.

Parallel Feeds. Figure 13.32 shows a number of parallel-feed systems. They frequently combine a number of radiators into subarrays, and the subarrays are then combined in series or in parallel to form sum and difference patterns.

Figure 13.32*a* shows a *matched corporate feed*, which is assembled from matched hybrids. The out-of-phase components of mismatch reflections from the aperture and of other unbalanced reflections are absorbed in the terminations. The in-phase and

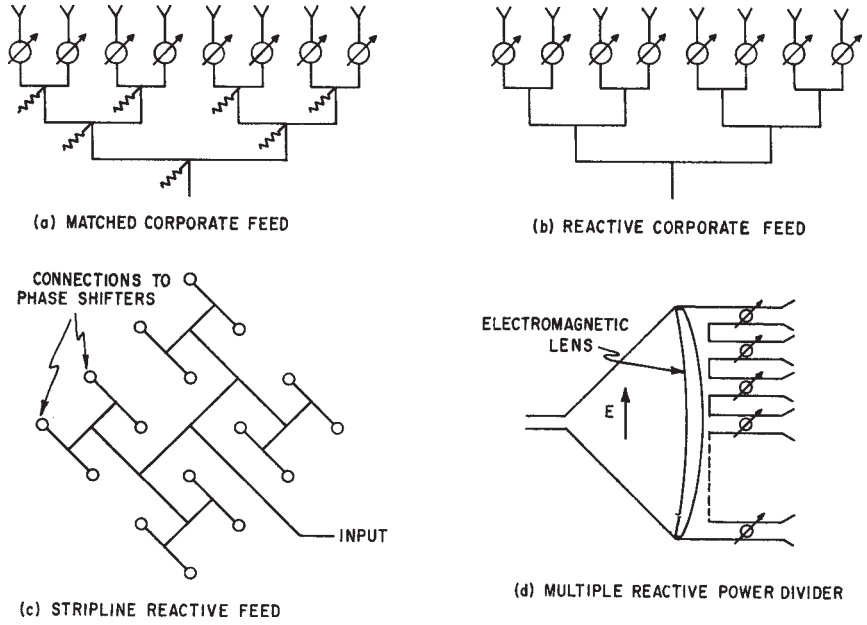


FIGURE 13.32 Parallel-feed networks

balanced components are returned to the input. To break up periodicity and reduce peak quantization lobes (Section 13.6), small additional fixed phase shifts may be introduced in the individual lines and compensated by corresponding readjustments of the phase shifters.

The reactive feed network shown in Figure 13.32*b* is simpler than the matched configuration. It has the disadvantage of not terminating unbalanced reflections that are likely to be at least partly reradiated and thus contribute to the sidelobes. A stripline power divider is shown in Figure 13.32*c*, and a constrained optical power divider using an electromagnetic lens is shown in Figure 13.32*d*. The lens may be omitted and the correction applied at the phase shifters. With nonreciprocal phase shifters, a fraction of the power reflected from the aperture will then be reradiated (as sidelobes) rather than returned to the input. The amplitude distribution across the horn is given by the waveguide mode. It is constant with an E-plane horn as shown.

Subarrays. The phased-array aperture may be divided into subarrays, all similar to simplify manufacturing and assembly. Beamforming now requires combining the subarrays for suitable sum and difference patterns. Figure 13.33*a* shows a method of combining opposite subarrays into their sums and differences. All sum channels are then added with proper weighting to obtain the desired amplitude distribution. The difference channels are treated similarly with independent amplitude weighting. This method may be extended to include combinations in the other plane.

Amplification on receiving or on both receiving and transmitting may be convenient at the subarray level. On receiving, the noise figure is established by the

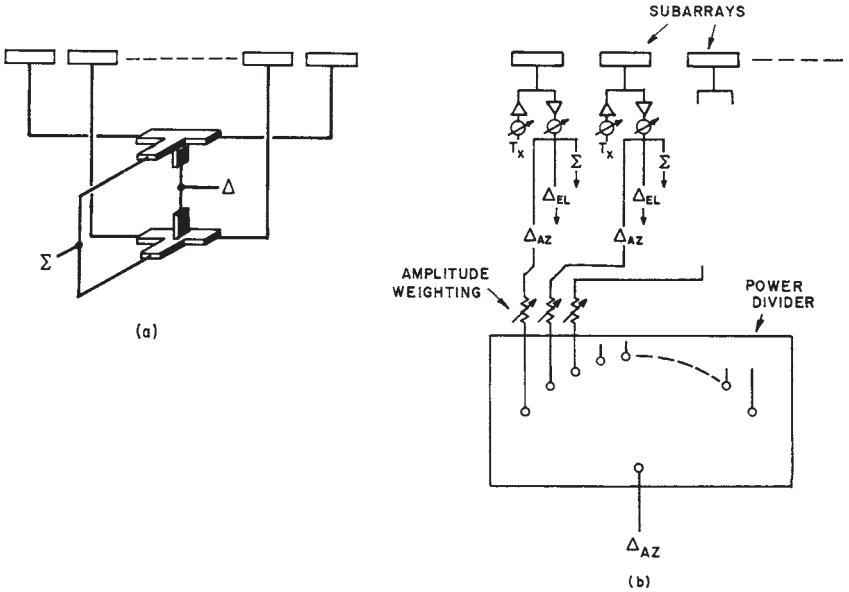


FIGURE 13.33 Combination of subarrays to form sum and difference channels: (*a*) combining opposite subarrays and (*b*) combining subarrays after amplification

preamplifier so that further processing may include lossy circuits. The receiving channel may be split three ways, as indicated in Figure 13.33*b*, into the sum and the elevation and azimuth difference channels. These are then weighted and summed with corresponding outputs from the other subarrays. On transmitting, all separate power amplifiers may be energized equally to give maximum power on the target. The addition of phase shifters shown at the subarray level simplifies the beam-steering computation, permitting all subarrays to receive identical steering commands. They may be replaced by time-delay circuits giving a wide instantaneous bandwidth (Section 13.7).

The simple method of providing TR (transmit-receive) switching can be obtained by combining two halves of a subarray with a four-port hybrid junction. The transmitter input into one port energizes both halves of the aperture, for example, in phase. The receiver, connected to the remaining port, then requires that the phase shifters of half the aperture give an additional phase shift of π during the receiving period. This is easily programmed.

Multiple receive beams may be formed by combining the subarrays after amplification in as many different ways as separate beams are required. The limitation is that the beams have to lie within the beamwidth of the subarray in order to avoid excessive grating lobes. A cluster of such simultaneous receiver beams requires a wider transmitter beam. This may be obtained efficiently from the same antenna with a gabled or spherical phase distribution.

With identical subarrays, a desired aperture amplitude taper (for low sidelobes) is applied with a granularity that depends on their size and shape. The resulting amplitude steps will cause quantization lobes (Section 13.6).

13.9 PHASE SHIFTERS

The three basic techniques for electronic beam steering are (1) frequency scanning, (2) beam switching, and (3) phase scanning with phase shifters. Of the three techniques, the use of phase shifters is by far the most popular, and considerable effort has gone into the development of a variety of phase shifters. Phase shifters can be separated into two categories: reciprocal and nonreciprocal. The reciprocal phase shifter is not directionally sensitive. That is, the phase shift in one direction (e.g., transmit) is the same as the phase shift in the opposite direction (e.g., receive). Therefore, if reciprocal phase shifters are used, it is not necessary to switch phase states between transmit and receive. With a nonreciprocal phase shifter, it is necessary to switch the phase shifter (i.e., change phase state) between transmit and receive. Typically, it takes a few microseconds to switch nonreciprocal ferrite phase shifters. During this time, the radar is unable to detect targets. For low pulse repetition frequency (PRF) radars [e.g., 200 to 500 pulses per second (pps)], this may not cause a problem. For example, if the PRF is 200 pps (or Hz), the time between pulses is 500 μs. If the switching time for the phase shifter is 10 μs, only 2% of the time is wasted and less than 1.0 nmi of minimum range is lost. On the other hand, if the PRF = 50 kHz, the time between pulses is 20 μs, and it would not be possible to tolerate 10 μs of dead time for the switching of phase shifters.

All diode phase shifters are reciprocal along with certain types of ferrite phase shifters. It is worth noting that, owing to losses associated with their magnetic properties, ferrite phase shifters are almost never used at frequencies below 3 GHz. Diode phase shifters, in contrast, improve as the frequency gets lower.

There are three basic types of phase shifters that typically compete for use in a phased array. They are (1) the diode phase shifter, (2) the nonreciprocal ferrite phase shifter, and (3) the reciprocal (dual-mode) ferrite phase shifter. Each has its strengths, and the choice of which to use is highly dependent on the radar requirements. Each will be discussed in turn. For solid-state systems, diodes are used and can be switched in a fraction of a microsecond.

Diode Phase Shifters.[87–90] Diode phase shifters are typically designed by using one of three techniques: (1) switched-line, (2) hybrid-coupled, and (3) loaded-line. These are shown diagrammatically in Figure 13.34. The switched-line technique simply switches in lengths of line in binary increments (e.g., 180°, 90°, and 45°) and requires a set of diodes for each bit. The diodes are used as switches to control which bits are activated to achieve a particular phase state.

The hybrid-coupled technique uses a microwave hybrid and effectively changes the distance at which the reflection takes place. This technique is usually used in binary increments, and an additional set of diodes is required for each phase state.

The diode phase shifters described above are limited in their ability to handle high peak power. Depending on their size and frequency, they are normally restricted to power levels of less than 1 kW. For higher power levels, the loaded-line technique is used. The diodes are used to switch in increments of capacitance and inductance that provide small changes in phase. Because the diodes are decoupled from the main transmission line, they need to handle only modest amounts of power in each diode. Very high power (i.e., kilowatt) configurations are possible. The technique does require many diodes, and the phase shifters are typically large and bulky as compared with the switched-line and hybrid-coupled techniques.

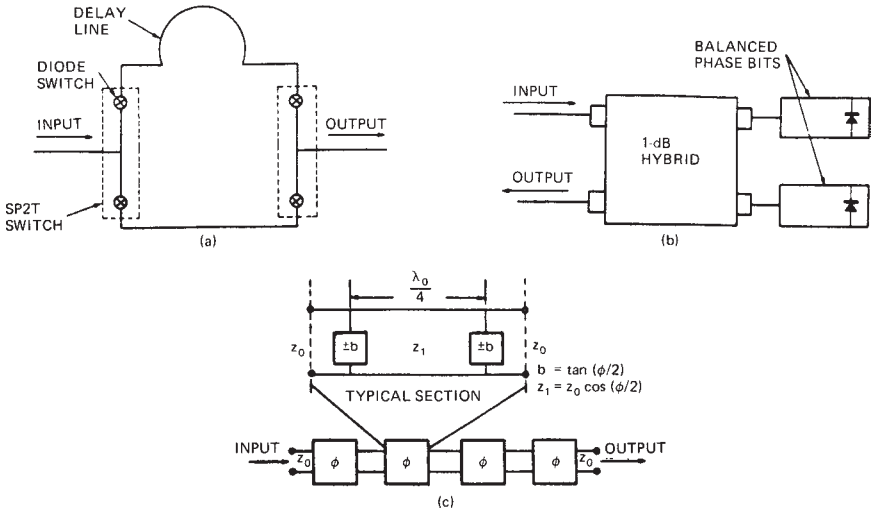


FIGURE 13.34 Diode phase-shifter configurations (*after L. Stark et al.*[87]): (*a*) switched-line phase bit, (*b*) hybrid-coupled phase bit, and (*c*) loaded-line phase bit

Diode phase shifters have the advantage of being small and light in weight (except for high-power devices). They are suitable for stripline, microstrip, and monolithic configurations. The main disadvantage of the diode phase shifter is that an additional set of diodes is normally required for each additional bit. As lower-sidelobe antennas are required, the number of bits increases. For low sidelobe antennas, 5, 6, or 7 bits may be required. As the number of bits increases, both cost and loss of the diode phase shifters are also increased. For active arrays, the phase shifter losses are not of significance because they occur prior to the power amplifier on transmit and after the low-noise amplifier on receive. This is not the case with most ferrite devices.

Ferrite Phase Shifters. Most ferrite phase shifters are nonreciprocal,[91–93] and their early versions used discrete lengths of ferrite (as shown in Figure 13.35) to implement each of the bits (180°, 90°, 45°, etc.). In this configuration, a current pulse is passed through each bit, and the ferrite toroid is saturated. When the current is removed, the ferrite toroid is said to be *latched* and retains its magnetization owing to its hysteresis properties. If the current is in a forward direction, the ferrite is latched with a particular phase (e.g., 180°). The ferrite maintains the phase until a current pulse in the opposite direction is applied. The ferrite phase shifter is then latched to the reference phase (0°). This change in phase with a change in current direction is due to the nonreciprocal nature of the device. As mentioned, early devices saturated each bit so that a ferrite toroid and electronic driver were required

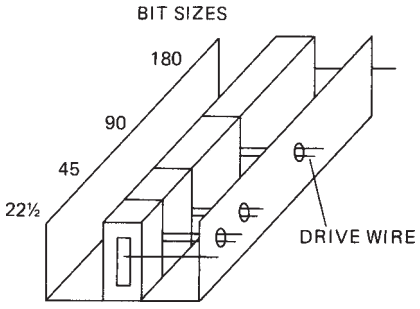


FIGURE 13.35 Digital ferrite phase shifter using toroids (*after L. Stark*[87])

to control each bit individually. Other phase shifters use a single toroid and a single driver.[94,95] In this configuration, the phase shifter is latched on a minor hysteresis loop by only partially magnetizing the ferrite. The distinct advantage of this technique is that any number of bits may be implemented while using only a single toroid. They have the advantage of low loss and relatively high power operation. Devices that handle up to 100 kW of peak power have been built. They are amenable to waveguide construction and are heavier and bulkier than comparable diode devices.

In summary, diodes and nonreciprocal ferrite phase shifters are viable competitors. At L band and lower, diode phase shifters are an obvious choice. At S band and higher, ferrites should continue to hold an edge in higher-power systems and where additional bits are needed for the low phase errors required for low-sidelobe antennas. These comments do not apply for the solid-state systems described below. Ferrite phase shifters are more temperature-sensitive than diodes, and the phase will change with a change in temperature. This can be controlled by keeping the temperature constant (within a few degrees) across the array. A more common technique is to sense the temperature at several locations in the array and then correct the phase commands to the phase shifters.

13.10 SOLID-STATE MODULES[96–98]

A solid-state module may be connected to every radiating element or to every subarray of a phased array antenna, forming what is sometimes called an *active aperture*. Applications range from ultrahigh frequency (UHF) for surveillance to X band and above for airborne systems.

A typical module is shown in simple schematic in Figure 13.36. It consists of a transmitter amplifier chain, a preamplifier for receiving, a shared phase shifter with driver, and circulators and/or switches to separate the transmit and receive paths (the gain around the loop must be less than unity to avoid oscillations).

Power amplifiers for transmitting at the element level would typically have a gain of 30 dB or more to compensate for the loss of power dividing in the beamformer. Transistors are capable of generating high average power but only relatively low peak power. High duty-cycle waveforms (10 to 20%) are therefore required to transmit enough energy efficiently. This lack in high peak power is the main disadvantage of the solid-state modules in phased array radars. To a great extent, it can be compensated for by using more pulse compression in the receiver and extra-wide bandwidth to counteract jamming, but at the cost of increased signal processing. An important advantage of transistors is their potential for wide bandwidth.

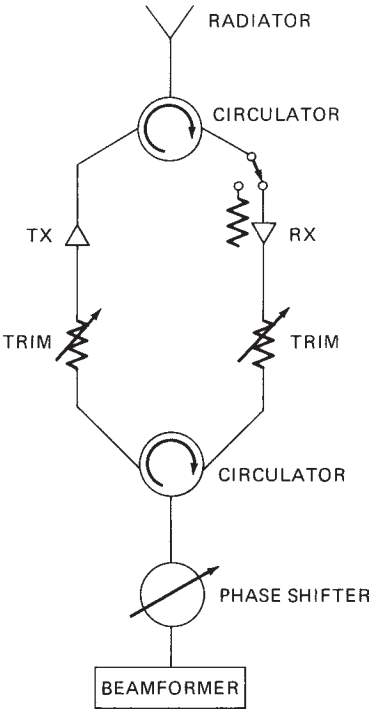


FIGURE 13.36 Typical solid-state module

The receiver requires an early gain of typically 10 to 20 dB to adequately establish a low noise figure and allow for phase shifting and beamforming losses. The module also receives interfering signals from all directions within the element pattern (not simply the antenna pattern), at all frequencies in the band. Tight tolerances between modules are needed to provide amplitude and phase tracking over the band for low-sidelobe performance. An electronically programmable gain adjustment may be helpful to correct module-to-module variations, allowing less demanding specifications. The trimmer would also provide a degree of freedom in aperture control for special situations. Because the noise figure has been established, the feed network may be split to give separate optimum aperture amplitude distributions for transmitting as well as for receiving on sum and difference channels. In an alternative configuration, the feed network could be designed for a constant amplitude aperture distribution to give highest transmit power on the target, and the receiver gain control could be used to provide an amplitude taper for the sum channel. Perhaps a secondary feed system could be added for the difference channels. Poirier[97] has analyzed this case, the effect on noise, and the degradation due to amplitude quantization steps.

The module phase shifter operates at low-signal power levels. Its insertion loss may be high because it is followed by amplification on transmit and preceded by amplification on receive. Diode phase shifters are, therefore, quite permissible even with many bits (e.g., 5, 6, or 7 bits for low sidelobes). Variations in insertion loss may be compensated for dynamically with the gain adjustment.

The circulator at the high-power side provides an impedance match to the power amplifier and may be adequate by itself for protecting the receiver. In Figure 13.36, the module is shown augmented by a switch that causes absorption of power reflected from antenna mismatches and also gives added protection to the receiver during transmission. If weight is an important consideration, as it would be in a space-based system, then the circulator could be replaced by diode switches requiring additional logic and driving circuitry.

13.11 MULTIPLE SIMULTANEOUS RECEIVE BEAMS

With the proliferation of ballistic missiles, future shipboard radar systems will probably need multi-mission capability, which includes both anti–air warfare (AAW) and ballistic missile defense (BMD). The BMD requirement to discriminate small radar cross section (RCS) re-entry vehicles (RVs) at long ranges requires that active array radars have a large power-aperture-gain/loss-noise temperature (PAG/LT), which is a measure of the sensitivity of the radar. Increasing the output power of a T/R module only increases the transmit power component of the PAG product. However, adding an additional T/R module to the array increases each of the components of the PAG: the transmit power, the receive aperture, and the transmit gain. Therefore, radars with a BMD function typically have a large number of elements, which results in a large aperture. Because a large antenna aperture produces a narrow antenna beamwidth, the volume search function will require scanning a large number of beam positions.

Naval ships need to operate in littoral (close to land) environments, as well as in the open ocean. The desire to operate closer to land requires shipboard radars to have the capability to reject high clutter returns. High clutter rejection is usually achieved through the use of pulse doppler waveforms in the first few elevation beam positions.

In addition, many of the radar waveforms may be designed to propagate out to the instrumented range of the radar. As a result, a major component of the dwell time at each beam position is the time required for the pulse to propagate to the instrumented range and back. This long propagation time, combined with the large number of beam positions due to a narrow antenna beamwidth and a pulse doppler waveform with a large number of pulses, results in unsatisfactory long search frame times when using a single beam radar.

If the active array radar is sized for long-range detection of ballistic missiles, the radar will most likely have sufficient power to detect other targets. A cluster of multiple simultaneous receive beams can be used to trade energy for a reduction in the volume search frame time. The use of multiple simultaneous receive beams will reduce the search frame time, since many beam positions are searched at the same time. When using multiple simultaneous receive beams, the transmit beam can be broadened by beam spoiling so that the 3-dB beamwidth is larger than that of a uniformly illuminated array with a linear phase front. Broadening the beam on transmit reduces the gain of the transmit beam, but this loss in gain may be necessary to produce the search frame time required. On receive, multiple beams are simultaneously formed in a cluster that covers the volume illuminated by the transmit beam. Each receive beam has the full gain and beamwidth of a single beam array with the same aperture dimensions. In the cluster of receive beams, each of the receive beams may be offset in various directions from the center of the spoiled transmit beam. An example of four simultaneous receive beams in a cluster is shown in Figure 13.37. To search the same volume covered by this four beam cluster with a single beam radar would take approximately four times as long since the transmit and receive beams would need to be sequentially scanned to each beam position. Increasing the number of beams in the receive cluster, while also increasing the transmit beam spoiling by an equivalent amount, reduces the search frame time.

There is another technique for generating multiple simultaneous beams that does not use beam broadening on transmit. In this technique, each pulse in the waveform is divided into as many segments as there are beam positions. A pulse segment is transmitted sequentially in each of the desired beam positions, one transmit pulse right after the other. After all the pulses have been transmitted, multiple simultaneous receive beams are formed with a receive beam located in each of the transmit beam positions. In this method, each of the transmit beams has the full array gain. In order to have similar performance to the beam-broadening technique, the transmit pulse segment in each beam position would be $1/N$ times as long as the transmit pulse in the beam-broadened technique, where N is the number of simultaneous receive beams. In either case, it is necessary to trade energy for time in order to achieve the desired search frame time.

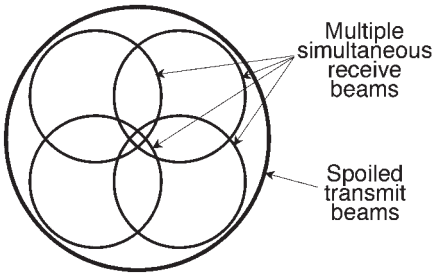


FIGURE 13.37 A spoiled transmit beam and a cluster of four simultaneous receive beams

A multiple simultaneous receive beam cluster can be formed in several different ways. If analog beamforming is used, the receive signals can be split N ways into N separate analog beamformers. Each analog beamformer is designed to produce one of the N offset beams in a beam cluster. Another way to generate multiple simultaneous receive beams is to use digital beamforming (DBF), which is discussed in more detail in the next section. DBF is usually preferred when a large number of beams are to be formed.

13.12 DIGITAL BEAMFORMING

Many phased array radars use analog beamforming. In an analog beamformer, the received signals from each element are combined at radio frequencies. At the output of the analog beamformer, centralized receivers downconvert the signal from radio frequency (RF) to an intermediate frequency (IF) and an analog-to-digital converter (ADC) is used to digitize the IF signal. In digital beamforming (DBF), the RF signals are digitized at either each individual element or each subarray. Once the signals have been digitized, various weighted combinations of the digitized signals can be used to form multiple simultaneous receive beams. Figure 13.38 shows array architectures for analog beamforming and DBF at the element and subarray level.

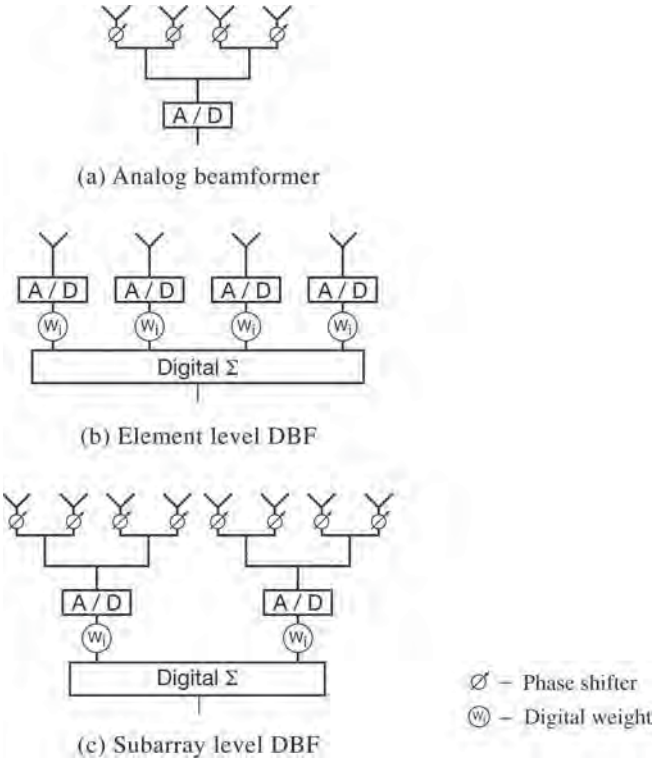


FIGURE 13.38 Array architectures for analog and digital beamforming

The benefits of DBF include improved dynamic range, faster search frame times if multiple simultaneous beams are used, and better control of amplitude and phase for adaptive nulling and lower sidelobe levels.

Shipboard radars had been designed to operate primarily in open ocean environments. However, the changing world will require shipboard radars to operate much closer to land. As a result, Navy radars must have significantly higher clutter rejection capabilities and larger dynamic ranges than existing radar systems. Recently, digital beamforming has been proposed for radar systems that must detect small RCS targets in severe clutter environments.[99]

The dynamic range determines the range of power levels that can be processed in the linear operating region of the receiver. The radar system must be able to process high power clutter returns without saturating the receiver. One reason DBF has been proposed for future radar is because DBF provides the radar system with a significantly higher dynamic range than can be achieved with analog beamforming, which has a centralized receiver. In a DBF architecture, there are N digital receivers, one at each of the N elements or subarrays. Therefore, the dynamic range of a DBF radar system has $10\log(N)$ higher dynamic range than an analog beamforming radar that utilizes the same receiver, provided the noise and distortion in each receiver are decorrelated among all receivers.[100]

DBF can be implemented at either the element level[101] or the subarray level. Element-level DBF requires a digital receiver, which consists of a downconverter and an ADC, at each element. Subarray-level DBF requires a digital receiver at each subarray. A fully digital array, which has DBF at the element level, allows for multiple, simultaneous independent beams. These independent beams can be steered in any direction. DBF at the subarray level can produce a multiple, simultaneous beam cluster located within the subarray pattern. DBF at the subarray level uses phase shifters at the element and an analog subarray beamformer, with a digital receiver located at the output of each subarray. Time delay steering can be achieved at the subarray level through the use of digital time delays. The multiple beam cluster is formed by digitally combining the subarrays to form simultaneous beams that are offset from one another. In practice, DBF is frequently implemented at the subarray level rather than the element level due to the size, weight, and high cost of digital receivers. Another practical advantage of subarray-level DBF is that less digital data, which must be processed to form the receive beams, is produced than for element-level DBF.

DBF can provide more efficient time-energy management than a single-beam analog beamforming array. Since DBF can produce multiple simultaneous beams, energy can be traded off for improved search frame times as was discussed in Section 13.11.

13.13 RADIATION PATTERN NULLING

Phased array antennas can be designed to have deterministic antenna patterns or adaptive antenna patterns. When no interfering signals are present, a phased array antenna with a deterministic antenna pattern will typically utilize phase weights that provide a linear phase front over the array face and amplitude weights that produce the desired sidelobe levels. The antenna performance can be characterized by parameters such as antenna beamwidth, gain, and peak and RMS sidelobe levels. Interfering radio frequency signals that are intentional, such as those produced by jammers, or unintentional, such as those produced by other radars or clutter, may significantly degrade the

performance of static-pattern phased array antennas. As a result, methods for placing nulls in the direction of the interference in the phased array radiation patterns have been the subject of much research.[102–105] Either deterministic radio frequency nulling or adaptive nulling can be used to place nulls in the antenna pattern in the direction of the interfering source.

If the location of the interfering source is stationary and the direction of the interferer is known, a deterministic RF null can be placed in the antenna pattern at that specific direction. Deterministic antenna patterns with nulls steered in specific directions can be achieved by modifying the weights at each element. These modified weights can be either the amplitude and phase of each element or only the phase of each element. Regardless of whether amplitude and phase nulling or phase-only nulling is used, the weights at each element are not time-varying for deterministic RF nulling. Array configurations for performing deterministic RF nulling using amplitude and phase weights and phase-only weights at each element are shown in Figure 13.39.

Deterministic RF nulling using amplitude and phase control at each element provides better performance than phase-only nulling. Deterministic RF nulling can be accomplished by using perturbation techniques to generate the weights at each element that will produce the antenna pattern with nulls.[102] With perturbation techniques, first the weights at each element are computed to generate the original antenna pattern without nulls. Then, the original weights are perturbed to produce nulls in the desired directions of the antenna pattern. The intent of the perturbation algorithm is to maximize the signal-to-interference-plus-noise ratio, while minimizing the deviation between the original element weights and the perturbed element weights.[104] Full amplitude and phase control at each element ensures that the perturbed antenna pattern has nulls in the desired directions with minimally perturbed weights.[102] However, phase-only control

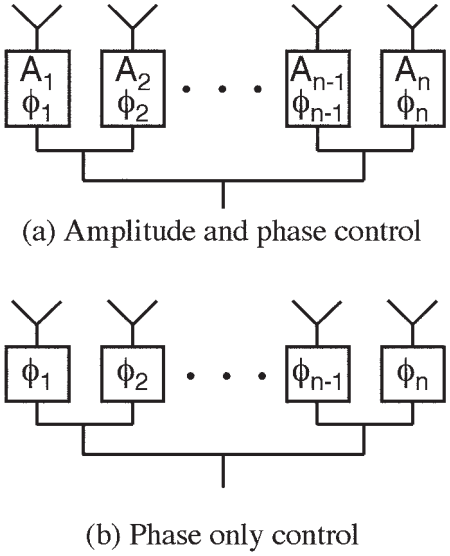


FIGURE 13.39 Array configurations for performing deterministic RF nulling: (*a*) amplitude and phase control at each element and (*b*) phase-only control at each element

at each element may result in situations where a solution does not exist.[102] An example of a deterministic antenna pattern with and without nulling is shown in Figure 13.40. The solid line in Figure 13.40 shows a deterministic antenna pattern at broadside for a 25-element linear array with $\lambda/2$ element spacing when no interference is present. Uniform illumination and no amplitude and phase errors were assumed. The dashed line in Figure 13.40 shows the antenna pattern for the same array when an interfering source is present at +21° and amplitude and phase weighting is used to steer a null in the pattern in the direction of the interfering source.

Radar systems can utilize deterministic RF nulling to place nulls in both the transmit and the receive antenna patterns. On transmit, nulls can be placed in the direction of high ground clutter to reduce the clutter power reflected back toward the antenna. Active arrays typically have both amplitude and phase control at each element, which allows either full amplitude and phase nulling or phase-only nulling to be used. However, when transmitting, it is desirable to have uniform illumination across the aperture to produce the highest aperture efficiency possible. Also, for active arrays, it is desirable to keep the high-power transmit amplifiers in saturation. Hence, when nulling on transmit, it is desirable to use phase-only nulling to maintain the aperture efficiency. On receive, amplitude weighting across the array is typically used to produce the desired sidelobe levels. Because this amplitude weighting produces a taper loss on receive, using full amplitude and phase control for nulling the radiation pattern is less of a concern than on transmit.

The incorporation of adaptive nulling into phased array antennas allows phased arrays to have adaptive antenna patterns on receive. Adaptive techniques are used to sense and automatically respond to a time-varying interference environment. The goal is to preserve the desired signal while simultaneously reducing unwanted interference. This real-time antenna pattern control is accomplished by adjusting the weights at each receiver to maximize signal-to-interference-plus-noise ratio. The amplitude and phase weight for the signal at each receiver is adaptively computed and implemented so that the desired signal combines coherently whereas the interference combines incoherently.

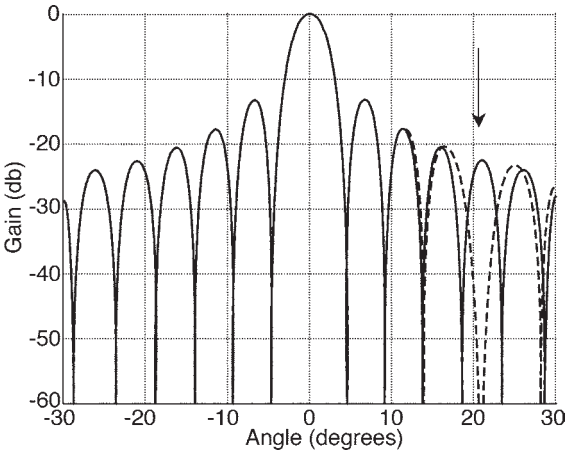


FIGURE 13.40 Deterministic antenna pattern with and without nulling for a 25-element linear array. Interfering source is located at +21 degrees.

Adaptive beamforming can be implemented by using auxiliary receivers in a sidelobe cancellation system or by using digital beamforming. In a DBF system, adaptive beamforming is accomplished either with the use of subarray receivers or with auxiliary receivers. If subarray receivers are used, the number of nulls that can be placed in the antenna pattern depends upon the number of degrees of freedom (DOF). For adaptive beamforming, the number of degrees of freedom is $N-1$, where N is the number of receivers in the DBF array.[100] If adaptive beamforming is implemented using digital beamforming, element-level DBF will have significantly better performance than subarray-level DBF, because subarray-level DBF has far fewer degrees of freedom. However, element-level DBF is usually cost prohibitive except for arrays with a small number of elements. Hence, subarray-level DBF is frequently pursued because it requires far fewer digital receivers. However, subarrays are not the best choice of array architectures for a given number of DOF in an adaptive array. [106] In subarray DBF, the main beam of each subarray must be aligned with the main-beam direction of the array. When an interfering source is present in the sidelobe region, the subarray-level DBF architecture does a poor job of sensing this interfering signal because the subarray gain is low in the sidelobe region.[106]

13.14 CALIBRATION OF ACTIVE PHASED ARRAY ANTENNAS

Active phased array antennas must be calibrated in order to ensure that the radiation pattern of the antenna meets the antenna performance specifications. Antenna characteristics such as sidelobe level and main-beam gain are typically specified, but these specifications can be translated into allowable amplitude and phase errors at each element.

Calibration can be divided into two categories: factory calibration and in-the-field calibration.[107] Factory calibration is only performed once, usually at the antenna test range of the phased array manufacturer. In-the-field calibration is performed at regular intervals throughout the life of the array, after it has been deployed. The antenna measurements required for factory calibration can be made in three different types of antenna ranges: near field, compact, or far field. For large ground-based or shipboard phased arrays, factory calibration is usually performed using a planar near-field antenna measurement facility.[108,109] An attractive feature of the near-field range is that the amplitude and phase of each element can be measured and a full hemispheric antenna pattern can be generated. Near-field ranges are suitable for measuring low-sidelobe, high-gain antennas.[110]

Planar near-field antenna measurements are made by scanning a probe in a plane parallel to the antenna under test and measuring the amplitude and phase at each probe location. In rectangular coordinates, the probe position is given by (x,y,z). During the scanning, the x and y coordinates of the probe are varied and the distance of the scanning plane from the antenna under test is kept constant at z_o. The value of z_o is typically a little more than three wavelengths. The phase front should be sampled with probe positions that satisfy the Nyquist sampling theorem, which yields a maximum sampling distance between points in the x-y plane of $\Delta x = \Delta y = \lambda/2$. In practice, the active array elements are turned on one at a time in coordination with the probe position. Some examples of existing systems that use near-field ranges to perform factory calibration are THAAD and SAMPSON.[108,109]

For an active array, both the phase and amplitude of each element must be measured in order to determine the correction constants for each array element in both transmit and receive mode. Full calibration of an active array is measurement intensive. It is common for phase shifters and attenuators to each have as many as 7 bits or 128 different states. In addition, active components should be characterized at several frequencies in the operating frequency band and over several temperatures. As a result, thousands of measurements are required to fully characterize the performance of each element. Not all states need to be measured for each element, but eliminating measurements at some of the states will cause the calibrated amplitude and phase errors to increase.[108]

Phase shifters usually exhibit small variations in amplitude over phase states, whereas attenuators exhibit large variations in-phase over gain states. Therefore, it is desirable to calibrate the attenuator first. The phase shifter can then be calibrated without changing the attenuator setting.[108] Once the characteristics of each T/R channel are measured, correction factors are calculated and stored for future use. The appropriate correction factors are applied to bring the antenna into calibration for a given mode of radar operation.

Once the active phased array is fielded, in-the-field calibration must be performed to keep the antenna within specifications. Active components drift over time and these aging effects require calibration to be performed in the field. A number of different techniques have been proposed for in-the-field calibration including mutual coupling,[111] near-field antennas,[109] and RF sampling.

Mutual-coupling techniques utilize the mutual coupling paths between adjacent array elements for transmission of calibration signals (see Figure 13.41*a*). In this technique, a signal is transmitted from an array element, and the nearest elements surrounding the transmit element are used to receive the transmitted calibration signal.[111] The signal received is compared to a stored reference obtained during factory tests. Only one element transmits at a given time, and this procedure is repeated until all the elements are measured. Calibration using mutual coupling can be interspersed with normal radar operating modes and requires only a small fraction of the available radar resources.

Another technique to perform in-the-field calibration uses radiators placed in the near field around the periphery of the array (see Figure 13.41*b*). These additional radiators are dedicated to the calibration of the active array. To calibrate the array in receive mode, test signals are routed to each of the radiators used for calibration.[109] The signal received from a T/R module is compared to a stored reference obtained during factory test.

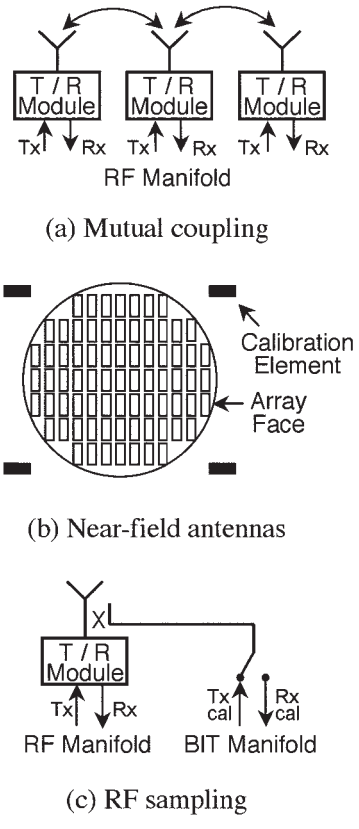


FIGURE 13.41 Active array calibration techniques: (*a*) mutual coupling, (*b*) near-field antennas, and (*c*) RF sampling

The calibration constants are then modified in order to restore the antenna so its calibration is as close as possible to the original factory calibration. This technique requires several additional transmit and receive hardware channels to be dedicated for in-the-field calibration.

RF sampling can also be used to calibrate an active phased array (see Figure 13.41*c*). The RF sampling technique uses built-in calibration circuitry in the array to inject a calibration signal between the T/R module and the radiator. This technique can be used to measure RF output power, receive gain, and phase and attenuation bit accuracy. These measured values are then compared with reference values. A major disadvantage of this technique is that it requires significant additional hardware, including a separate built-in test (BIT) manifold. Also, the calibration path does not include the radiator.

13.15 PHASED ARRAY SYSTEMS

A number of phased array radar systems have been built,[112] and a representative selection is briefly reviewed below.

Lockheed Martin Corporation Phased Array Radars. (*Courtesy of Lockheed Martin Corporation*)

AN/SPY-1.[113,114] This S-band phased array radar is part of the Navy Aegis weapon system and was developed by RCA, now Lockheed Martin. It has four phased array apertures to give unobstructed hemispherical coverage (Figure 13.1). In its early configuration, it used a simple feed system with, on receive, 68 subarrays, each containing 64 waveguide-type radiators for a total of 4352 elements. On transmit, subarrays were combined in pairs, and 32 such pairs gave a transmit aperture of 4096 radiators. The phase shifters fed directly into waveguide radiators, had 5 bits, and were of the nonreciprocal, latching garnet configuration. A later version was designed for low sidelobes. The subarray size had to be reduced to 2 elements to avoid quantization lobes, and similarly, the phase shifter had to be refined by driving it with 7-bit accuracy. The resulting phased array has an aperture with a constrained-feed structure and 4350 waveguide-type radiators. Monopulse sum and difference receive patterns and the transmit pattern are separately optimized.

COBRA. The COunter Battery RAdar (COBRA) C-Band solid -state phased array radar (see Figure 13.42) locates rocket launchers, artillery batteries, mortars, shell impact points, and jamming directions. The antenna, which is mechanically trained, has 2720 radiating elements, each fed by a solid state T/R module to provide wide electronic scan coverage. It is produced by Euro-Art GmbH, a consortium of Lockheed Martin MS2 (Moorestown, NJ), EADS Deutschland (Unterschleissheim, Germany), Thales Air Defence (Bagneux, France), and Thales UK, Limited (Crawley, West Sussex, England), for the Ministries of Defence of France, Germany, and the United Kingdom.

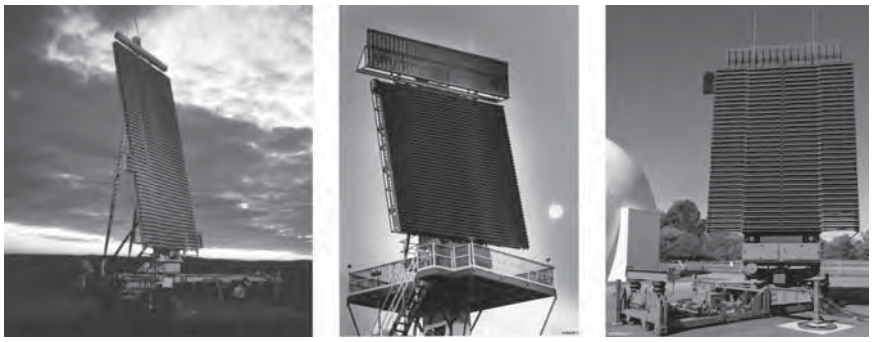
Volume Search Radar. The S-band Volume Search Radar (VSR) phased array (see Figure 13.43) provides search and AAW capability to the Dual Band Radar System for DD(X). It has 2688 radiating elements with a T/R module per element. Although the active radiating aperture is approximately circular, the corners have been filled in

FIGURE 13.42 COBRA phased array mounted on radar system vehicle (*Courtesy of Lockheed Martin Corporation*)

with terminated elements to form a rectangular aperture for purposes of minimizing antenna radar cross section. All active components are accessible from the rear for ease of maintenance aboard ship.

Ground-Based Radars (AN/TPS-59, AN/FPS-117, and AN/TPS-77). Lockheed Martin's long-range, solid-state 3D radar systems (see Figure 13.44) operate at L band between 1.2–1.4 GHz. These systems provide position data for en-route tracking of commercial aircraft, air surveillance/air defense, navigational assistance, tactical control, and tactical ballistic missile (TBM) defense. The AN/TPS-59 was the world's first totally solid-state antenna/array. The AN/FPS-117 was developed shortly thereafter and is typically installed and operated at a fixed site, but it can also be configured on a trailer for deployment. The AN/TPS-77 is a mobile/transportable series of solid-state radar systems. Each system incorporates primary surveillance radar, secondary surveillance radar (IFF), command and control display consoles, communications equipment, various options, and transport configurations customized for each customer's specific mission.

FIGURE 13.43 Front (left) and rear views of Volume Search Radar S-Band engineering development model active phased array mounted in near-field test facility during assembly. Rear view shows detail of easily replaceable lowest replaceable units. (*Courtesy of Lockheed Martin Corporation*)



AN/TPS-59 AN/FPS-117 AN/TPS-77

FIGURE 13.44 Ground-based radar system configurations (*Courtesy of Lockheed Martin Corporation*)

The main feature of these systems is the planar phased array, which has modular construction with distributed all-solid-state electronics, including RF power generation and electronic steering in the elevation axis. Antenna rotation for the system is at 5, 6 or 10, 12 revolutions per minute (rpm), providing 360° azimuth coverage while electronically step-scanning in elevation. It provides full performance over a nominal 20° in the elevation plane to an altitude of 100,000 feet and up to 60° in elevation during TBM tracking. The antenna/array incorporates monopulse radar techniques. These techniques allow accurate target angular position, using three data channels—sum, delta azimuth (AZ), and delta elevation (EL)—within the array and the data-processing group.

The data-processing group applies moving target indicator (MTI) and doppler processing to the receive signal to produce subclutter visibility for target detection. Selection of the operating frequency is dependent upon the radar's mode of operation and the frequency options selected by the operator. The antenna/array shown in Figure 13.44 is constructed of either 54 (AN/TPS-59), 44 (AN/FPS-117), or 34 (AN/TPS-77) identical horizontal passive row-feed networks, one above another on 6.6-inch centers. Mounted directly behind each row feed is a dedicated row transmitter and receiver. The row transmitters and receivers are similar for each row. They perform all the RF functions, including RF power generation, phase shifting to steer the antenna beam in elevation, and receive preamplification to provide a low-noise figure. The RF power is generated by direct amplification through modular solid-state power amplifiers in the row transmitters. The energy from these power modules is combined within the row transmitter, which drives the row feed. The RF energy from the row feeds is combined in space to form the radiated beam. Unlike a tube transmitter, which operates at high peak power and low duty factor, the power modules operate at low peak power and high duty factors (10% to 20%). Long-duration pulses are transmitted and then compressed during receive to obtain the required range resolution. A performance margin is built in to the radar system to allow for field degradation of the distributed array components between scheduled maintenance periods. Since array component failures induce only a gradual degradation, maintenance can be performed on a planned basis. Online system sensitivity and performance level is checked and displayed by the performance monitoring and fault isolation (MFI) function.

These long-range, three-dimensional air-surveillance radars provide an off-the-shelf solution to primary air-surveillance needs. They have been designed for remote operation in harsh environments, and with their established logistics support, built-in performance monitoring and fault isolation, high reliability, and worldwide investment, these systems have an inherent low cost of ownership.

Northrop Grumman Corporation Phased Array Radars. (*Courtesy of Northrop Grumman Corporation*)

AWACS (Airborne Warning and Control System).[13,14,115] The Airborne Warning and Control System (AWACS) provides all-weather long-range surveillance, command, control, and communications. The phased array antenna is a 26 ft by 4.5 ft elliptically shaped, ultra-low sidelobe array, consisting of 28 slotted waveguides with over 4000 slots (see Figure 13.45). Mounted atop the aircraft fuselage in a rotating dome, the S-band antenna scans mechanically over 360° in azimuth, and scans electronically in elevation through 28 reciprocal beam-steering phase shifters. In addition, 28 nonreciprocal beam offset phase shifters provide offset of the receive beam from the transmit beam during elevation scanning to compensate for time delay between transmit and receive of long-range aircraft returns. The Radar System Improvement Program (RSIP) provides the most significant upgrade to the AWACS radar since its development in the early 1970s. RSIP introduces new waveforms and processing, providing improvements in detection range, angular accuracy, as well as range and angular resolution.

AN/TPS-78. The AN/TPS-78 is an update to the AN/TPS-70/75 S-band, three-dimensional tactical radar family developed by the Northrop Grumman Corporation (Figure 13.46). The AN/TPS-78 planar array shown in Figure 13.46 exhibits very low

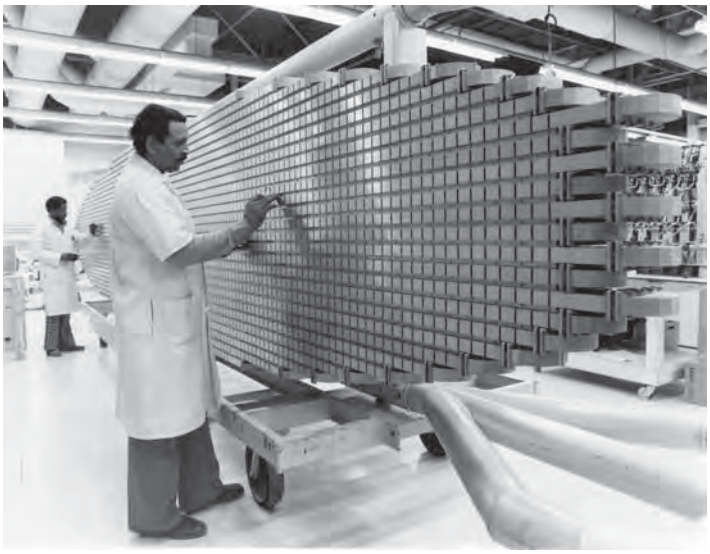


FIGURE 13.45 AWACS antenna (*Courtesy of Northrop Grumman Corporation*)

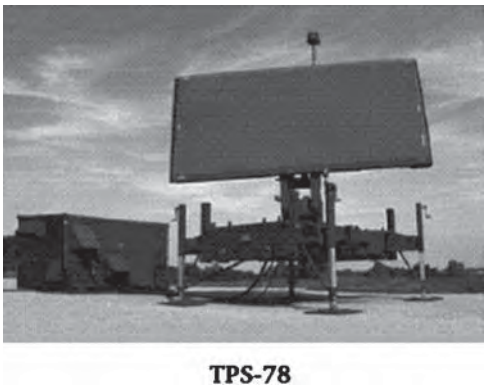


FIGURE 13.46 AN/TPS-78 antenna (*Courtesy of Northrop Grumman Corporation*)

azimuth sidelobes. The array is a slotted waveguide design of 36 sticks and 94 horizontally polarized slots per stick. The array is fed from a solid-state transmitter and transmits a single fan beam in elevation optimized for maximum gain at low altitudes. On receive, the array output is combined into six simultaneous narrow elevation beams that cover from 0 to 20°. Each receive beam is processed in a digital receiver. Accurate height information is extracted using amplitude monopulse techniques by comparing receive energy in adjacent beams. A narrow azimuth beamwidth on transmit and receive and stacked receive beams provide good radar angle resolution. Azimuth coverage of 360° is achieved by rotating the array at a nominal surveillance rate.

A unique feature of the AN/TPS-78 array is that the array has an integral low sidelobe IFF array embedded in the array face. This not only provides excellent registration between radar and IFF returns, but also supports fast setup and teardown time for transporting the radar in its mobile configuration.

Joint STARS Antenna. The Joint Surveillance Target Attack Radar System (Joint STARS) provides long-range airborne surveillance and tracking of stationary and moving ground targets (see Figure 13.47). The phased array antenna for this system is a 24-ft wide by 2-ft high planar X-band slotted array. It is mounted inside a radome underneath the fuselage of an E-8A aircraft and can be mechanically rotated ±100° in elevation for operation looking out either side of the aircraft. The system operates in a synthetic aperture radar (SAR) mapping mode as well as in a three-port azimuth interferometer ground moving target indication (GMTI) target detection and tracking mode. Use of the three-port aperture on receive enables a clutter suppression interferometry technique that allows the operator to detect, accurately locate, and track slow moving ground targets in the presence of competing ground clutter.

Raytheon Company Phased Array Radars. (*Courtesy of Raytheon Company*)

PATRIOT.[116] The PATRIOT is a multifunction phased array radar system developed by Raytheon for the Army in the form of a lens using an optical feed, as shown previously in Figure 13.30. Sum and difference patterns are separately optimized with a monopulse feed. The aperture is round and contains approximately 5000 elements.

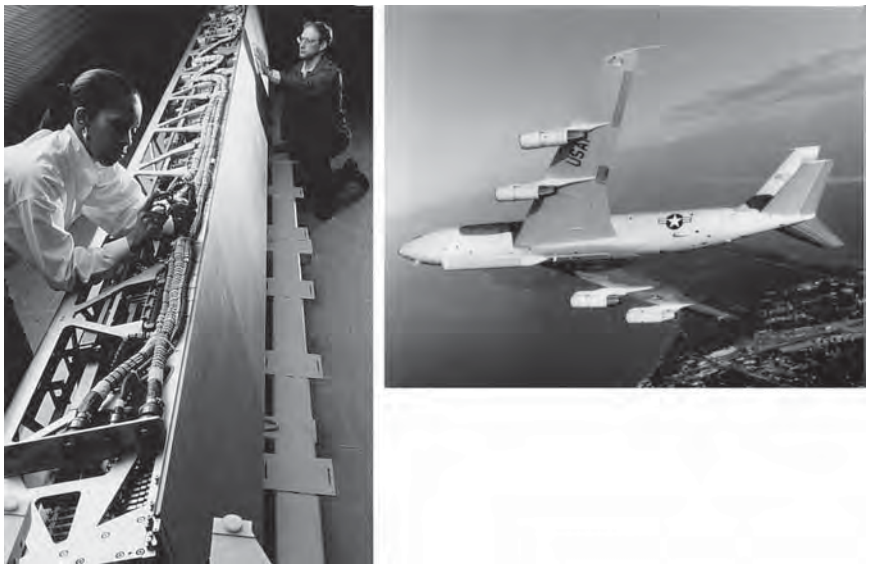


FIGURE 13.47 Joint STARS antenna and aircraft (*Courtesy of Northrop Grumman Corporation*)

It uses 4-bit flux-driven nonreciprocal ferrite phase shifters and waveguide-type radiators at both apertures. The antenna is shown in Figure 13.48. It is mounted on a vehicle and folds flat for transportation.

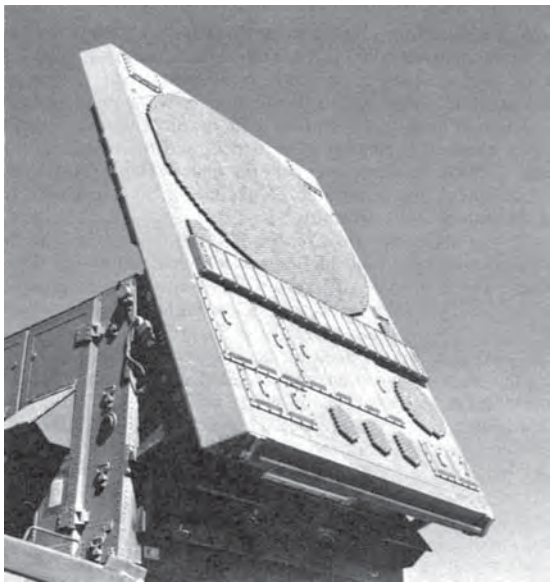


FIGURE 13.48 PATRIOT multifunction phased array (*Courtesy of Raytheon Company*)

Terminal High Altitude Area Defense (THAAD). THAAD is a ground-based system designed to destroy the theater ballistic missile threats to troops, military assets, and allied territories. THAAD consists of a hit-to-kill missile, radar, launcher, and Battle Management/Command, Control, and Communications (BMC3) system.

The THAAD radar (see Figure 13.49) is an X-band, phased array, solid-state radar capable of search, threat detection, classification, and precision tracking at long ranges. The radar design delivers high power output and beam/waveform agility in order to support the long-range functional requirements of the THAAD mission. The THAAD element components work in concert to detect, assign, and destroy incoming short- to medium-range ballistic missiles by providing the firing battery with target acquisition, interceptor support, and intercept assessment functions. These functions include surveillance, THAAD missile track, in-flight data uplink/downlink, target classification/typing/identification, and intercept assessment.

The single face phased-array radar is 12.5 m long, mounted on a Heavy Expanded Mobile Tactical Truck and capable of being transported in a C-130 Hercules aircraft. The radar is nonrotating and has a 120º field of view. The radar uses its 9.2 m^2 full field-of-view phased array aperture to acquire theater missile threats at a range of up to 1,000 km. This radar uses 25,344 X-band solid-state transmit/receive modules in the phased array.

X-Band Radar (XBR). The nine-story-high XBR is the world's largest X-band radar, weighing four million pounds (see Figure 13.50). The SBX platform that it sits on stands more than 250 ft tall and displaces more than 50,000 tons. It consists of a semi-submersible oil production platform, topped with the XBR. XBR is the primary payload on the semi-submersible platform supporting the Ground-Based Midcourse Defense (GMD) phase of the Missile Defense Agency (MDA) Ballistic Missile Defense System (BMDS). SBX's floating platform, a modified oil-drilling vessel, was designed for

FIGURE 13.49 THAAD radar (*Courtesy of Raytheon Company*)

FIGURE 13.50 SBX—over 250 ft tall and displacing more than 50,000 tons (*Courtesy of Raytheon Company*)

exceptional stability in high winds and storms. Measuring 240 ft wide and 390 ft long, the vessel includes a power plant, bridge and control rooms, living quarters, storage areas, and enough floor space and infrastructure to support the X-band radar.

The X-band radar itself, which sits on top of the floating platform, is the largest, most sophisticated phased array, electro-mechanically steered X-band radar in the world. It consists of thousands of elements driven by transmit/receive (T/R) modules. In the X-band radar, they will provide the full fire control sensor functions for the Ground-Based Midcourse Defense system, including search, acquisition, tracking, discrimination, and kill assessment.

SPY-3. The AN/SPY-3 Multi-Function Radar (MFR) is the U.S. Navy's first shipboard active phased array multifunction radar. It is an X-band active phased array radar designed to meet all horizon search and fire control requirements for the Navy. MFR is designed to detect low-observable anti-ship cruise missile (ASCM) threats and support fire-control illumination requirements for multiple missiles. MFR combines the functions provided by several separate radars aboard Navy combatant ships and also supports new ship-design requirements for reduced radar cross section and significantly reduced manning (no operators).

The radar performs such functions as horizon search, limited above-the-horizon search, and fire control track and illumination. One of the most significant design features of the radar is to provide automatic detection, tracking, and illumination of low-altitude threat missiles in the adverse environmental conditions routinely found in coastal waters.

SPY-3 uses three fixed-face arrays, each containing around 5,000 transmit/receive (T/R) elements (see Figure 13.51). These elements are connected to T/R integrated multi-channel modules, each of which drives eight elements. Individual modules are designed to slide into the array structure and provide a high-conductivity thermal path to the cooling-array manifold without having any connection to the T/R module itself.

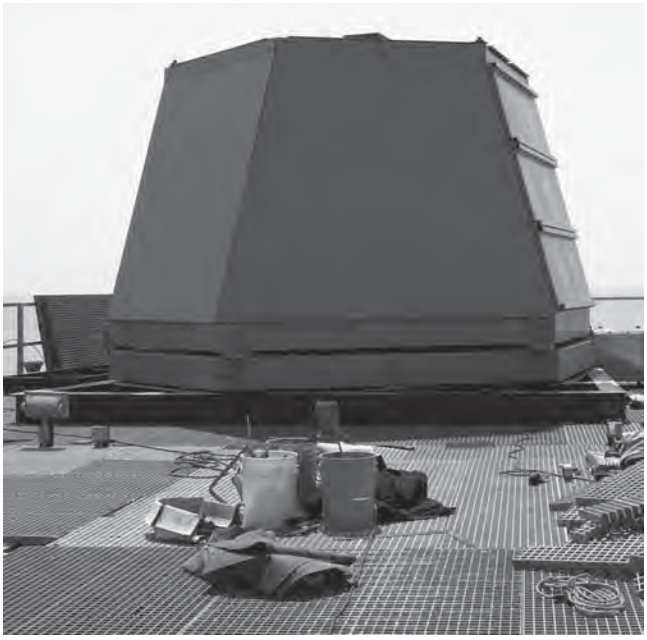


FIGURE 13.51 AN/SPY-3 Multi-Function Radar (*Courtesy of Raytheon Company*)

REFERENCES

1. A. Blondel, "Improvements in or relating to radiator systems for wireless telegraphy," Belgian Patent 163,516, 1902; British Patent 11, 427, 1903.
2. S. G. Brown, "Improvements in wireless telegraphy," British Patent 14,449, 1899.
3. R. M. Foster, "Directive diagrams of antenna arrays," *Bell System Tech. J*., vol. 5, pp. 292–307, April 1926.
4. R. C. Hansen, *Microwave Scanning Antennas*, Vols. I, II, and III, New York: Academic Press, 1964.
5. N. Amitay, R. C. Pecina, and C. P. Wu, "Radiation properties of large planar arrays," *Bell Teleph. Syst. Monog*., 5047, February 1965.
6. A. A. Oliner and G. H. Knittel, *Phased Array Antennas*, Norwood, MA: Artech House, 1972.
7. L. I. Stark, "Microwave theory of phased array antennas—A Review," *Proc. IEEE*, vol. 62, pp. 1661–1701, December 1974.
8. R. J. Mailloux, "Phased array theory and technology," *Proc. IEEE*, vol. 70, March 1982.
9. A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, *The Handbook of Antenna Design*, Vol. 2, London: Peter Peregrinus, Ltd., 1983.
10. R. C. Johnson and H. Jasik (eds.), *Antenna Engineering Handbook*, 2nd Ed., New York: McGraw-Hill Book Company, 1984.
11. E. Brookner, "Phased array radars," *Sci. Am*., vol. 252, pp. 94–102, February 1985.
12. H. P. Steyskal, "Phased arrays 1985 symposium," in *RADC Rept. TR*-85-171, Rome Air Development Center, Bedford, MA, August 1985.
13. E. Brookner, "A review of array radars," *Microwave J*., vol. 24, pp. 25–114, October 1981.
14. E. Brookner, "Radar of the 80's and beyond," *IEEE Electro 84*, May 1984.

15. E. Brookner, "Array radars: An update," *Microwave J*., vol. 30, pt. I, pp. 117–138, February 1987; pt. II, pp. 167–174, March 1987.
16. W. A. Harmening, "A laser-based, near-field probe position measurement system," *Microwave J*., pp. 91–102, October 1983.
17. R. J. Mailloux, Chapter 21 in *Antenna Engineering Handbook*, R. C. Johnson and H. Jasik (eds.), 2nd ed., New York: McGraw-Hill Book Company, 1984.
18. G. V. Borgiotti, Chapter 11 in *The Handbook of Antenna Design*, A. W. Rudge, K. Milne, A. D. Olver and P. Knight (eds.), vol. 2, London: Peter Peregrinus, Ltd., 1983.
19. K. G. Schroeder, "Near optimum beam patterns for phase monopulse arrays," *Microwaves*, pp. 18–27, March 1963.
20. J. Frank, "Phased array antenna development," Johns Hopkins University, Appl. Phys. Lab. Rept. TG 882, pp. 114–117, March 1967.
21. W. E. Scharfman and G. August, "Pattern measurements of phased arrayed antennas by focussing into the near zone," in "Phased Array Antennas," A. A. Oliner and G. H. Knittel (eds.), Norwood, MA: Artech House, 1972, pp. 344–349.
22. D. K. Alexander and R. P. Gray, Jr., "Computer-aided fault determination for an advanced phased array antenna," presented at Proc. Antenna Application Symp., Allerton, IL., September 1979.
23. J. Ronen and R. H. Clarke, "Monitoring techniques for phased-array antennas," *IEEE Trans*., vol. AP-33, pp. 1313–1327, December 1985.
24. R. H. Collings, "Current sheet antenna," U.S. Patent 3,680,136, 1972.
25. A. A. Oliner and R. G. Malech, Chapters 2–4 in *Microwave Scanning Antennas*, R. C. Hansen, vol. III, New York: Academic Press, 1964.
26. L. I. Parad and R. W. Kreutel, "Mutual effects between circularly polarized elements," in *Symp. USAF Antenna Res. Develop., Antenna Arrays Sec., Abstr*., University of Illinois, Urbana, 1962.
27. M. I. Skolnik, *Introduction to Radar Systems*, 2nd Ed., New York: McGraw-Hill Book Company, 1980, pp. 504–506.
28. J. M. Howell, "Limited scan antennas," presented at IEEE AP-5 Int. Symp., 1972.
29. J. S. Ajioka, "Frequency-scan antennas," *Antenna Engineering Handbook*, Chap. 19, R. C. Johnson and H. Jasik (eds.), 2nd Ed., New York: McGraw-Hill Book Company, 1984.
30. W. F. Gabriel, Guest editor, special issue on adaptive antennas, *IEEE Trans*., vol. AP-24, September 1976.
31. J. R. Forrest, Guest editor, special issue on phased arrays, *Proc. IEE* (*London*), vol. 127, pt. F, August 1980.
32. H. Steyskal, "Digital beamforming antennas—an introduction," *Microwave J*., vol. 30, pp. 107–124, January 1987.
33. J. L. Allen, "A theoretical limitation on the formation of lossless multiple beams in linear arrays," *IRE Trans*., vol. AP-9, pp. 350–352, July 1961.
34. J. Blass, "The multidirectional antenna: A new approach to stacked beams," in *IRE Int. Conv. Rec*., vol. 8 pt. 1, 1960, pp. 48–50.
35. H. Gent, The bootlace aerial, *Royal Radar Estab. J*. (*U.K.*), pp. 47–57, October 1957.
36. W. Rotman and R. F. Turner, "Wide angle microwave lens for line source application," *IEEE Trans*., vol. AP-11, pp. 623–632, November 1963.
37. S. A. Schelkunoff, "A mathematical theory of linear arrays," *Bell, Syst. Tech. J*., vol. 22, pp. 80–107, January 1943.
38. J. F. Ramsay, J. P. Thompson, and W. D. White, "Polarization tracking of antennas," presented at IRE Int. Conv., Session 8, Antennas I, 1962.
39. P. M. Woodward, A method of calculating the field over a planar aperture required to produce a given polar diagram, *J. IEE* (*London*), vol. 93, pt. 3A, pp. 1554–1558, 1946.
40. R. C. Hansen, "Aperture theory," in *Microwave Scanning Antennas*, vol. I, New York: Academic Press, pp. 18–21.
41. J. F. Ramsay, "Lambda functions describe antenna diffraction pattern," *Microwaves*, pp. 70–107, June 1967.

42. A. Ksienski, "Equivalence between continuous and discrete radiating arrays," *Can. J. Phys*., vol. 39, pp. 35–349, 1961.
43. W. H. Von Aulock, "Properties of phased array," *IRE Trans*. Vol. AP-9, pp. 1715–1727, October 1960.
44. E. D. Sharp," Triangular arrangement of planar-array elements that reduces number needed," *IRE Trans*., vol. AP-9, pp. 126–129, March 1961.
45. J. L. Allen et al., "Phased array radar studies," MIT Lincoln Lab. Tech. Rept. 381, March 1965.
46. J. L. Allen and B. L. Diamond, "Mutual coupling in array antennas," MIT Lincoln Lab. Tech. Rept. 424, October 1966.
47. R. W. Bickmore, "Note on effective aperture of electrically scanned arrays," *IRE Trans*., vol. AP-6, pp. 194–196, April 1958.
48. B. L. Diamond, Chapter 3 in *Mutual Coupling in Array Antennas*, J. L. Allen and B. L. Diamond (eds.), MIT Lincoln Lab. Tech. Rept. 424, pt. III, October 1966.
49. J. Frank, "Phased array antenna development," *Johns Hopkins University, Appl. Phys. Lab. Rept. TG* 882, March 1967.
50. P. W. Hannan, "Element-gain paradox for a phased-array antenna," *IEEE Trans*., vol. AP-12, pp. 423–43, July 1964.
51. W. Wasylkiwskyj and W. K. Kahn, "Element patterns and active reflection coefficient in uniform phased arrays," *IEEE Trans*., vol. AP-22, March 1974.
52. W. Wasylkiwskyj and W. K. Kahn, "Element pattern bounds in uniform phased arrays." *IEEE Trans*., vol. AP-25, September 1977.
53. W. K. Kahn, "Impedance-match and element-pattern constraints for finite arrays," *IEEE Trans*., vol. AP-25, November 1977.
54. R. E. Willey, "Space tapering of linear and planar arrays," *IRE Trans*., vol. AP-10, pp. 369–377, July 1962.
55. H. A. Wheeler, "Simple relations derived from a phased-array antenna made of an infinite current sheet," *IEEE Trans*., vol. AP-13, pp. 506–514, July 1965.
56. J. L. Allen, "On array element impedance variation with spacing," *IEEE Trans*., vol. AP-12, p. 371, May 1964.
57. P. W. Hannan, "The ultimate decay of mutual coupling in a planar array antenna," *IEEE Trans*., vol. AP-14, pp. 246–248, March 1966.
58. T. T. Debski and P. W. Hannan, "Complex mutual coupling measured in a large phased array antenna," *Microwave J*., pp. 93–96, June 1965.
59. P. W. Hannan and M. A. Balfour, "Simulation of a phased-array antenna in a waveguide," *IEEE Trans*., vol. AP-13, pp. 342–353, May 1965.
60. B. L. Diamond, "Small Arrays—Their analysis and their use for the design of array elements," in *Phased Array Antennas*, A. A. Oliner and G. H. Knittel (eds.), Norwood, MA: Artech House,1972.
61. C. E. Grove, D. J. Martin, and C. Pepe, "Active impedance effects in low sidelobe and ultra wideband phased arrays," *Proc. Phased Arrays Symp*., 1985, pp. 187–206.
62. G. E. Evans and H. E. Schrank, "Low sidelobe radar antennas," *Microwave J*., pp. 109–117, July 1983.
63. G. E. Evans and S. G. Winchell, "A wide band, ultralow sidelobe antenna," presented at Antenna Applications Symp., Allerton, IL, September 1979.
64. S. G. Winchell and D. Davis, "Near field blockage of an ultralow sidelobe antenna," *IEEE Trans*., vol. AP-28, pp. 451–459, July 1980.
65. D. K. Barton and H. R. Ward, *Handbook of Radar Measurement*, Englewood Cliffs, NJ: Prentice-Hall, 1969, pp. 242–338.
66. J. F. Ramsey, "Lambda functions describe antenna/diffraction patterns," *Microwaves*, p. 60, June 1967.
67. W. M. Yarnall, "Twenty-seven design aids for antennas, propagation effects and systems planning," *Microwaves*, pp. 47–73, May 1965.

68. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, January 1978.
69. T. T. Taylor, "Design of line source antennas for narrow beamwidth and low sidelobes," *IEEE Trans*., vol. AP-3, pp. 16–28, 1955.
70. T. T. Taylor, "Design of circular apertures for narrow beamwidth and low sidelobes," *IEEE Trans*., vol. AP-8, pp. 17–22, 1960.
71. E. T. Bayliss, "Design of monopulse antenna difference patterns with low sidelobes," *Bell Syst. Tech. J*., pp. 623–650, May–June 1968.
72. D. K. Barton and H. R. Ward, *Handbook of Radar Measurement*, Englewood Cliffs, NJ: Prentice-Hall, 1969, pp. 256–266.
73. J. L. Allen, "The theory of array antennas," MIT Lincoln Lab. Rept. 323, July 1963.
74. J. Ruze, "Physical limitations on antennas," MIT Res. Lab. Electron. Tech. Rept. 248, October 30, 1952.
75. T. C. Cheston, "Effect of random errors on sidelobes of phased arrays," *IEEE APS Newsletter—Antenna Designer's Notebook*, pp. 20–21, April 1985.
76. J. Frank and J. Ruze, "Steering increments for antisymmetrically phased arrays," *IEEE Trans*., vol. AP-15, pp. 820–821, November 1967.
77. J. Brown (Private communication), 1951.
78. C. J. Miller, "Minimizing the effects of phase quantization errors in an electronically scanned array," in *Proc. Symp. Electronically Scanned Array Techniques and Applications*, RADC-TDR-64-225, vol. 1, 1964, pp. 17–38.
79. J. Frank, "Bandwidth criteria for phased array antennas," in *Phased Array Antennas*, A. A. Oliner and G. H. Knittel, Norwood, MA: Artech House, 1972, pp. 243–253.
80. W. B. Adams, "Phased array radar performance with wideband signals," *AES Conv. Rec*., November 1967, pp. 257–271.
81. C. B. Sharp and R. B. Crane, "Optimization of linear arrays for broadband signals," *IEEE Trans*., vol. AP-14, pp. 422–427, July 1966.
82. C. Rothenberg and L. Schwartzman, "Phased array signal bandwidth," IN *IEEE Int. Symp. Antennas Propag. Dig*., December 1969, pp. 116–123.
83. W. Rotman and R. F. Turner, "Wide angle lens for line source applications," *IEEE Trans*., vol. AP-11, pp. 623–632, 1963.
84. J. Blass, "The multi-directional antenna: a new approach to stacked beams," in *Proc. IRE Conv*., vol. 8, pt. I, 1960, pp. 48–51.
85. R. F. Kinsey and A. Horvath, "Transient response of center-series fed array antennas," in *Phased Array Antennas*, A. A. Olinerand and G. H. Knittel (eds.), Norwood, MA: Artech House, 1972, pp. 261–271.
86. A. R. Lopez, "Monopulse networks for series feeding an array antenna," in *IEEE Int. Symp. Antennas Propag. Dig*., 1967.
87. L. Stark, R. W. Burns, and W. P. Clark, Chapter 12 in *Radar Handbook*, M. I. Skolnik (ed.), 1st Ed., New York: McGraw-Hill Book Company, 1970.
88. J. F. White, *Semiconductor Control*, Norwood, MA: Artech House, 1977.
89. W. J. Ince, "Recent advances in diode and phase shifter technology for phased array radars," pts. I and II, *Microwave J*., vol. 15, no. 9, pp. 36–46, and no. 10, pp. 31–36, 1972.
90. J. F. White, "Diode phase shifters for array antennas," *IEEE Trans*., vol. MTT-22, pp. 658–674, June 1974.
91. M. A. Fruchaft and L. M. Silber, "Use of Microwave ferrite toroids to eliminate external magnets and reduce switching power," *Proc. IRE*, vol. 46, p. 1538, August 1958.
92. J. Frank, C. A. Shipley, and J. H. Kuck, "Latching ferrite phase shifter for phased arrays," *Microwave J*., pp. 97–102, March 1967.
93. W. J. Ince and D. H. Temme, "Phase shifters and time delay elements," in *Advances in Microwaves*, vol. 4, New York: Academic Press, 1969.

94. L. R. Whicker and C. W. Young, "The evolution of ferrite control components," *Microwave J*., vol. 2, no. 11, pp. 33–37, 1978.
95. J. DiBartolo, W. J. Ince, and D. H. Temme, "A solid state 'flux drive' control circuit for latching-ferrite-phase shifter applications," *Microwave J*., vol. 15, pp. 59–64, September 1972.
96. R. A. Pucel (ed.), *Monolithic Microwave Integrated Circuits*, New York: IEEE Press, 1985.
97. J. L. Poirier, "An analysis of simplified feed architectures for MMIC T/R module arrays," *Rome Air Development Center Rept. RADC-TR*-86-236 (AD A185474), February 1987.
98. W. H. Perkins and T. A. Midford, "MMIC technology: better performance at affordable cost," *Microwave J*., vol. 31, pp. 135–143, April 1988.
99. B. Cantrell, J. de Graaf, F. Willwerth, G. Meurer, L. Leibowitz, C. Parris, and R. Stapleton, "Development of a digital array radar," *IEEE AEES Systems Magazine*, pp. 22–27, March 2002.
100. H. Steyskal, "Digital beamforming antennas: an introduction," *Microwave J*., pp. 107–124, January 1987.
101. R. C. Hansen, *Phased Array Antennas*, New York: John Wiley & Sons, 1998.
102. H. Steyskal, R. A. Shore, and R. L. Haupt, "Methods for null control and their effects on the radiation pattern," *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 404–409, March 1986.
103. W. F. Gabriel, "Adaptive processing array systems," *Proc. of the IEEE*, vol. 80, no. 1, pp. 152–161, January 1992.
104. S. P. Applebaum, "Adaptive arrays," *IEEE Trans. Antennas and Propagation*, vol. AP-24, no. 5, pp. 585–598, September 1976.
105 R. J. Mailloux, *Phased Array Antenna Handbook*, Norwood, MA: Artech House, 2005.
106. M. Zatman, "Digitization requirements for digital radar arrays," in *Proceedings of the 2001 IEEE Radar Conference*, May 1–3, 2001, pp. 163–168.
107. G. H. C. van Werkhoven and A. K. Golshayan, "Calibration aspects of the APAR antenna unit," in *IEEE International Conference on Phased Array Systems and Technology*, May 21–25, 2000, pp. 425–428.
108. J. K. Mulcahey and M. G. Sarcione, "Calibration and diagnostics of the THAAD solid state phased array in a planar nearfield facility," *IEEE International Symposium on Phased Array Systems and Technology*, 1996, pp. 322–326, October 15–18, 1996.
109. M. Scott, "Sampson MFR active phased array antenna," in *IEEE International Symposium on Phased Array Systems and Technology*, October 14–17, 2003, pp. 119–123.
110. D. Slater, *Near-field Antenna Measurements*, Norwood, MA: Artech House, 1991.
111. H. M. Aumann, A. J. Fenn, and F. G. Willwerth, "Phased array antenna calibration and pattern prediction using mutual coupling measurements," *IEEE Trans. Antennas and Prop*., vol. 37, no.7, pp. 844–850, July 1989.
112. E. Brookner, "Phased arrays and radars – past, present, and future," *Microwave J*., vol. 49, pp. 24–46, January 2006.
113. R. M. Scudder and W. H. Sheppard, "AN/SPY-1 phased array antenna," *Microwave J*., vol. 17, pp. 51–55, May 1974.
114. R. L. Britton, T. W. Kimbrell, C. E. Caldwell, and G. C. Rose, "AN/SPY-1 planned improvements," in *Conf. Rec. Eascon '82*, September 1982, pp. 379–386.
115. B. Walsh, "An eagle in the sky," *Countermeasures—The Military Electron. Mag*., pp. 30–63, July 1976.
116. D. R. Carey and W. Evans, "The PATRIOT radar in tactical air defense," *Microwave J*., vol. 31, pp. 325–332, May 1988.

Chapter 14
Radar Cross Section

Eugene F. Knott

Tomorrow's Research

14.1 INTRODUCTION

A radar detects or tracks a target, and sometimes can classify it, only because there is an echo signal. It is, therefore, critical in the design and operation of radars to be able to quantify or otherwise describe the echo, especially in terms of such target characteristics as size, shape, and orientation. For that purpose, the target is ascribed an effective area called the *radar cross section*, or RCS. The RCS is the projected area of a metal sphere that would return the same echo signal as the target, had the sphere been substituted for it.

Unlike the echo of the sphere, however, which is independent of the viewing angle, the echoes of all but the simplest targets vary significantly with orientation. As will be shown later, this variation can be quite rapid, especially for targets many wavelengths in size.

The echo characteristics depend in strong measure on the size and nature of the target surfaces exposed to the radar beam. The variation is small for electrically small targets (targets less than a wavelength or so in size) because the incident wavelength is too long to resolve target details. On the other hand, the flat, singly curved and doubly curved surfaces of electrically large targets all give rise to different echo characteristics. Reentrant structures such as jet engine intakes and exhausts generally have large echoes, and even the trailing edges of airfoils can be significant echo sources.

The radar cross sections of simple bodies can be computed exactly by a solution of the wave equation in a coordinate system for which a constant coordinate coincides with the surface of the body. The exact solution requires that the electric and magnetic fields just inside and just outside the surface satisfy certain conditions that depend on the electromagnetic properties of the material of which the body is made.

While these solutions constitute interesting academic exercises and can, with some study, reveal the nature of the scattering mechanisms that come into play, there are no known tactical targets that fit the solutions. Thus, exact solutions of the wave equation are, at best, guidelines for gauging other (approximate) methods of computing scattered fields.

An alternative approach is the solution of the integral equations governing the distribution of induced fields on target surfaces. The most useful approach to a solution is known as the *method of moments*, in which the integral equations are reduced to a system of linear homogeneous equations. The attraction of this method is that the surface profile of the body is unrestricted, allowing the computation of the scattering

from truly tactical objects. Another is that ordinary methods of solution (matrix inversion and gaussian elimination, for example) may be employed to effect a solution. This method is limited by computer memory and execution time, however, to objects a few dozen wavelengths in size at best.

Alternatives to these exact solutions are several approximate methods that may be applied with reasonable accuracy to electrically large target features. They include the theories of geometrical and physical optics, the geometrical and physical theories of diffraction, and the method of equivalent currents. These approximations are discussed in Section 14.3. Other approximate methods not discussed here are explored in detail in some of the references listed at the end of this chapter.

The practical engineer cannot rely entirely on predictions and computations, and must eventually measure the echo characteristics of some targets. This may be done by using full-scale test objects or scale models thereof. Small targets often may be measured indoors, but large targets must usually be measured on an outdoor test range. The characteristics of typical indoor and outdoor test facilities are described in Section 14.4.

Control of the echo characteristics of some targets is of vital tactical importance, namely *stealth*. There are only two practical ways of reducing the echo: through shaping and radar absorbers. *Shaping* is the selection or design of surface profiles so that little or no energy is reflected back toward the radar. Because target contours are difficult to change once the target has become a production item, shaping is best implemented in the concept definition stage before production decisions have been made. Radar-absorbing materials actually soak up radar energy, also reducing the energy reflected back to the radar. However, the application of absorbers is always expensive, whether gauged in terms of nonrecurring engineering costs, lifetime maintenance, or reduced mission capabilities. The two methods of echo control are discussed in Section 14.5, and a collection of four stealthy platforms is surveyed there.

Seven Basic Echo Mechanisms. Figure 14.1 illustrates seven basic echo sources that might be found on a typical airborne target. All depend in varying degree on the

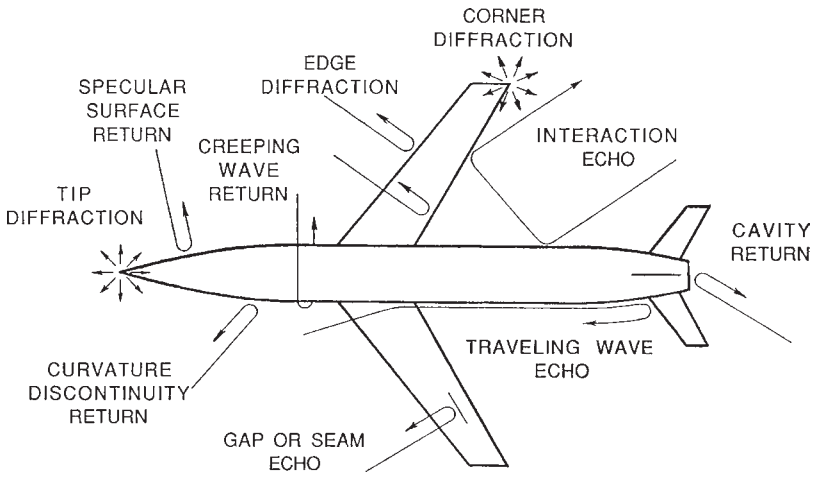


FIGURE 14.1 Examples of seven basic echo-source mechanisms (*after E. F. Knott*[1] *© American Institute of Aeronautics and Astronautics, 1992*)

target aspect angle as seen from the radar. Some are dominant scattering mechanisms whereas others are weak. Not all are significant on other kinds of platforms, such as warships or military ground vehicles. We briefly discuss the seven in descending order of significance.

Reentrant Structures. The only reentrant structure appearing on the hypothetical missile in Figure 14.1 is its exhaust duct at the rear, but jet intake ducts behave much the same way. The echoes from cavities such as intake ducts, exhaust ducts, and cockpits are large and tend to persist over aspect angles as wide as 45° or 60°. This is because most of the internal duct surfaces (i.e., compressor stages and turbine faces) are metallic, and any radar wave that finds its way into the structure will likely find its way back out toward the radar. This is also true of internal reflections from within cockpit canopies.

Specular Scatterers. A specular scatterer is any target surface that is oriented perpendicular to the line of sight to the radar. Flat surfaces offer particularly large echoes in the specular direction, but the echoes drop off sharply away from that direction. The specular echoes from singly and doubly curved surfaces (cylindrical and spheroidal surfaces) are somewhat lower than those from flat surfaces, but are more persistent with changes in aspect angle.

Traveling-wave Echoes. When the angle of incidence is a small grazing angle off the surface, a surface traveling wave can be induced. The surface wave tends to build up toward the rear of the body and is usually reflected back toward the front by any discontinuity at the rear. Traveling-wave echoes at low grazing angles are nearly as significant as specular echoes at normal incidence.

Tip, Edge, and Corner Diffraction. Scattering from tips, edges, and corners is less significant than specular echoes and thus are worrisome to the designer only when most other sources of echo have been suppressed. The echoes from tips and corners are localized and tend to increase with the square of the wavelength, not the size of any surface feature. Thus, they become progressively less important as the radar frequency rises.

Surface Discontinuities. Most airframes have slots or gaps where control surfaces meet the stationary airframe. Slots, gaps, and even rivet heads can reflect energy back to the radar. Because these tend to be small effects, it is not easy to isolate and characterize them.

Creeping Waves. A creeping wave is one that gets bound to a smooth, shaded surface, is guided around the rear of a smooth body, and is then launched back to the radar when it reappears at the shadow boundary on the opposite side. As shown in the next section, the creeping wave causes the echoes from small spheres to vary with sphere size. The mechanism can also be present for other smooth bodies, such as the generic missile depicted in Figure 14.1. The creeping wave mechanism is never a significant one for military and civilian targets.

Interactions. Relatively strong echoes can occur when a pair of target surfaces are oriented for a favorable bounce from one surface to another and then back to the radar, as in the interaction between the fuselage and the trailing edge of the right wing shown

in Figure 14.1. Similar interactions occur for ship targets when bulkheads, railings, masts, and other topside features become mirrored in the mean sea surface.

Not all of these mechanisms are revealed in the characteristics of the selection of simple and complex targets, as shown in the next section.

14.2 THE CONCEPT OF ECHO POWER

Definition of RCS. An object exposed to an electromagnetic wave disperses incident energy in all directions. This spatial distribution of energy is called *scattering*, and the object itself is often called a *scatterer*. The energy scattered back to the source of the wave (called *backscattering*) constitutes the *radar echo* of the object. The intensity of the echo is described explicitly by the radar cross section of the object, for which the acronym *RCS* has been generally recognized. Early papers on the subject called it the *echo area* or the *effective area*, terms still found occasionally in contemporary technical literature.

The formal definition of radar cross section is

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \frac{|E_s|^2}{|E_0|^2} \tag{14.1}$$

where E_0 is the electric-field strength of the incident wave impinging on the target and E_s is the electric-field strength of the scattered wave at the radar. The derivation of the expression assumes that a target extracts power from an incident wave and then radiates that power uniformly in all directions. Although the vast majority of targets do *not* scatter energy uniformly in all directions, the definition assumes that they do. This permits one to calculate the scattered power density on the surface of a large sphere of radius R centered on the scattering object. R is typically taken to be the range from the radar to the target.

The symbol σ has been widely accepted as the designation for the RCS of an object, although this was not so at first.[2,3] The RCS is the projected area of a metal sphere that is large compared with the wavelength and that, if substituted for the object, would scatter identically the same power back to the radar. However, the RCS of all but the simplest scatterers fluctuates greatly with the orientation of the object, so the notion of an equivalent sphere is not very useful.

The limiting process in Eq. 14.1 is not always an absolute requirement. In both measurement and analysis, the radar receiver and transmitter are usually taken to be in the far field of the target (discussed in Section 14.4), and at that distance, the scattered field E_s decays inversely with the distance R. Thus, the R^2 term in the numerator of Eq. 14.1 is canceled by an identical but implicit R^2 term in the denominator. Consequently, the dependence of the RCS on R, and the need to form the limit, usually disappears.

The radar cross section is, therefore, a comparison of the scattered power density at the receiver with the incident power density at the target. An equally valid definition of the RCS results when the electric-field strengths in Eq. 14.1 are replaced with the incident and scattered magnetic-field strengths. It is often necessary to measure or calculate the power scattered in some other direction than back to the transmitter, a *bistatic* situation. A bistatic RCS may be defined for this case as well as for backscattering, provided it is understood that the distance R is measured from the target to the receiver.

Forward scattering is a special case of bistatic scattering in which the bistatic angle is 180° whence the direction of interest is along the shadow zone behind the target.

The shadow itself can be regarded as the sum of two fields of nearly equal strength but 180° out of phase. One is the incident field, and the other is the scattered field. The formation of the shadow implies that the forward scattering is large, which is indeed the case. The fields behind the target are hardly ever precisely zero, however, because some energy usually leaks into the shadow zone via diffraction from the sides of the target.

Examples of RCS Characteristics. The discussion of radar cross section characteristics will first consider simple targets, of which the sphere is a classic example. This will be followed by complex objects, of which an aircraft is a good example.

Simple Objects. Because of its pure radial symmetry, the perfectly conducting sphere is the simplest of all three-dimensional scatterers. Despite the simplicity of its geometrical surface, however, and the invariance of its echo with orientation, the RCS of the sphere varies considerably with electrical size. The exact solution for the scattering by a conducting sphere is known as the Mie series[4] and is shown in Figure 14.2.

Note that the chart is split roughly into three regions. In the *Rayleigh region* below $ka < 1$, the RCS rises with the fourth power of the sphere radius, an echo dependence characteristic of electrically small bodies, whether spherical or not. In this region, the incident wave cannot accurately resolve body length-to-width variations. In the optics region above $ka > 10$, optics formulas for predicting the RCS of the metal body generally work reasonably well. Sandwiched between the Raleigh region below and the optics region above is the resonance region, where two or more mechanisms may combine in and out of phase with each other to produce the undulations in the RCS.

In the case of the sphere, the undulations in the resonance region are due to two distinct contributions to the echo, one a *specular reflection* from the front of the sphere and the other a *creeping wave* that circles around its shadowed side. The two go in and out of phase as the sphere grows larger because the difference in the lengths of their electrical path from source to receiver increases continuously with increasing *ka*. The undulations become weaker with increasing *ka* because the creeping wave loses more energy the longer the electrical path around the shadowed side.

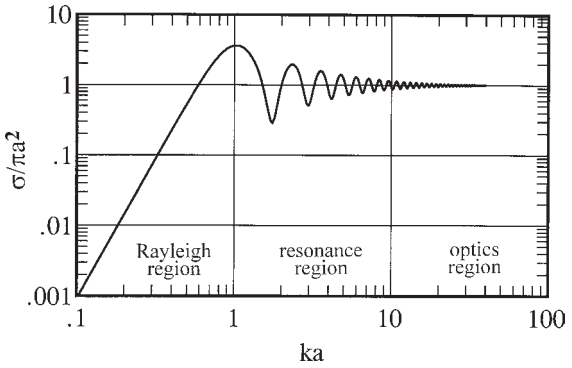


FIGURE 14.2 Radar cross section of a perfectly conducting sphere normalized to the optics value πa^2. The parameter $ka = 2\pi a/\lambda$ is the circumference of the sphere expressed in wavelengths.

If only the specular reflection is significant, the optics-region RCS of the perfectly conducting sphere is simply

$$\sigma = \pi a^2 \tag{14.2}$$

where a is the radius of the sphere. But the RCS of permeable (dielectric) bodies is more complicated than this because energy can enter the body and rattle around inside before coming back out. An example is the dielectric sphere whose RCS is plotted in Figure 14.3. Because the dielectric material is slightly lossy, as indicated by the nonzero imaginary component of the index of refraction, the RCS of the sphere decays gradually with increasing electrical size. Atlas et al. went even further by comparing the measured and theoretical RCS of Plexiglas spheres in their attempts to understand the scattering by hailstones.[6]

The RCS of very slender dielectric bodies does not exhibit this complexity, however, because the sources of reflection (front and back sides of a dielectric cylinder, for example) are too close to each other to be resolvable by the incident wave. An example is the broadside RCS of a thin string shown in Figure 14.4. The string was angled 45° across the test zone of a large indoor test chamber, and the RCS was measured as a function of frequency for four transmit-receive polarization combinations. Only the copolarized VV and cross-polarized VH traces are shown because the HH and HV measurements closely track the VV and VH data. The measured data are the rapidly varying traces and were fitted statistically to the smoothly varying traces representing the exact solution of the two-dimensional wave equation for dielectric cylinders.

The string diameter was 0.012 inch and its illuminated length was estimated to be about 37 ft. Based on a mean separation between the measured VV and VH data of 10.7 dB, the effective dielectric constant of the string was estimated to be $\varepsilon_r = 2.646$. This may be the first time that RCS measurements were ever used to estimate the dielectric constant of a string. Strings are of interest at RCS test facilities because they sometimes are used as "invisible" target supports.

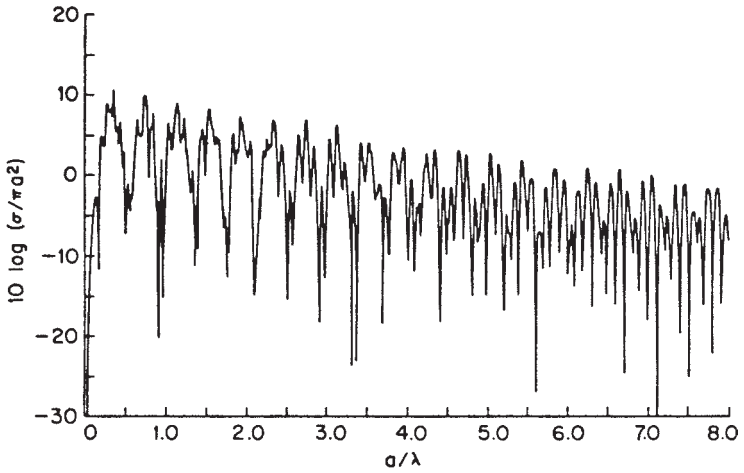


FIGURE 14.3 RCS of a lossy dielectric sphere with $n = 2.5 + i0.01$ (*after J. Rheinstein*[5] *© IEEE 1968*)

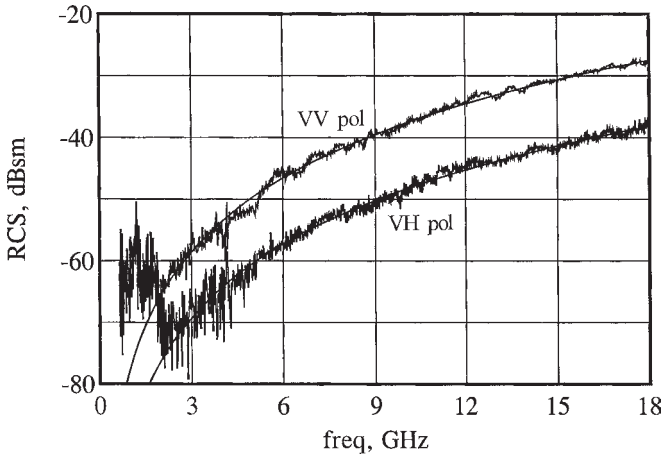


FIGURE 14.4 Measured and predicted broadside RCS of a string stretched across the test zone of an indoor test chamber at 45° angle (© *1999 Horizon House.*[7] *Reprinted with permission.*)

The behavior of short wire dipoles is markedly different from that of long dielectric strings. As shown in Figure 14.5, the broadside echo of a metal wire exhibits resonances at odd multiples of a half wavelength, with plateaus of nearly constant return between the resonant peaks. These plateaus rise with increasing dipole length and become less distinct as the dipole becomes thicker and longer. They eventually disappear when the dipole becomes fat enough and long enough.

The RCS can rise to significant levels at end-on aspects, as well as in the broadside regions, of bodies both fat and thin. These near-end-on echoes are attributable to surface traveling waves that radiate power in the backward direction. An example is the ogive, a spindle-shaped object formed by rotating an arc of a circle about its chord.

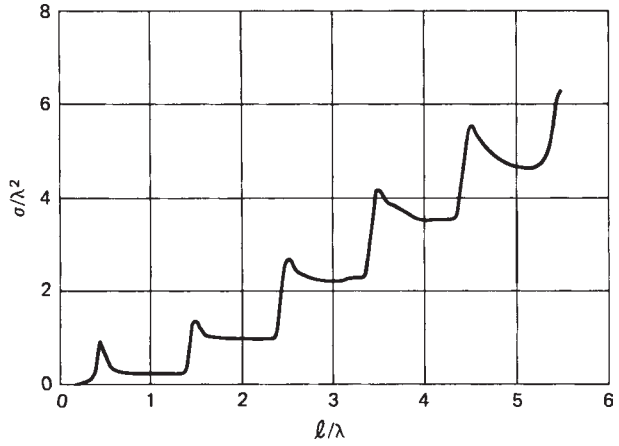


FIGURE 14.5 Measured broadside returns of a thin dipole (*Courtesy of University of Michigan Radiation Laboratory*[8])

Figure 14.6 is the RCS pattern of a 39λ long, 15° half-angle ogive recorded for horizontal polarization (incident electric field in the plane of the ogive axis and the line of sight). The large lobe at the right side of the pattern is a specular echo in the broadside sector, and the sequence of peaks at the left side is the contribution of the surface traveling wave near end-on incidence. Note that the RCS is extremely small (not measurable in this case) at precisely end-on incidence. Theoretical predictions in the end-on region closely match the measured pattern for this particular body.

A metal plate is a more elementary structure than the ogive shown in Figure 14.6, but its RCS pattern is no less complex. Sample patterns are reproduced in Figure 14.7 for four different incident and received polarization combinations. The traces for HH (dots) and HV (short dashes) are each shifted down by 5 dB for clarity, whereas those for the VV (long dashes) and VH (solid) are "as is" (not shifted). The plate was rotated about a vertical axis parallel to one edge of the plate, and the incident or received polarization was either parallel to (V) or perpendicular to (H) that axis, respectively.

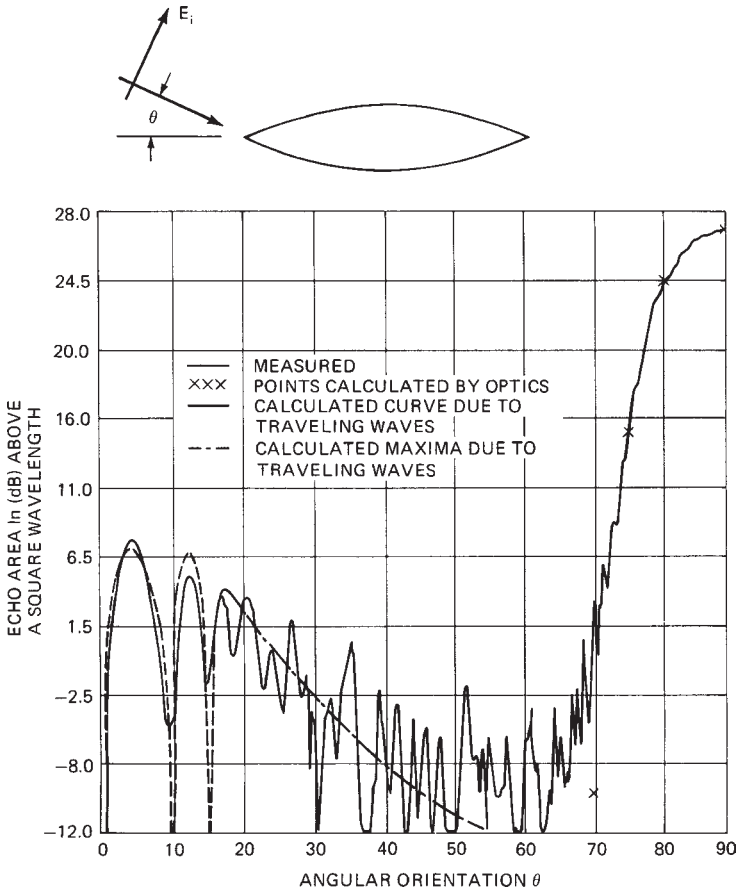


FIGURE 14.6 Measured RCS pattern of a 39λ, 15° half-angle metal ogive (*after L. Peters*[9] © *IEEE 1958*)

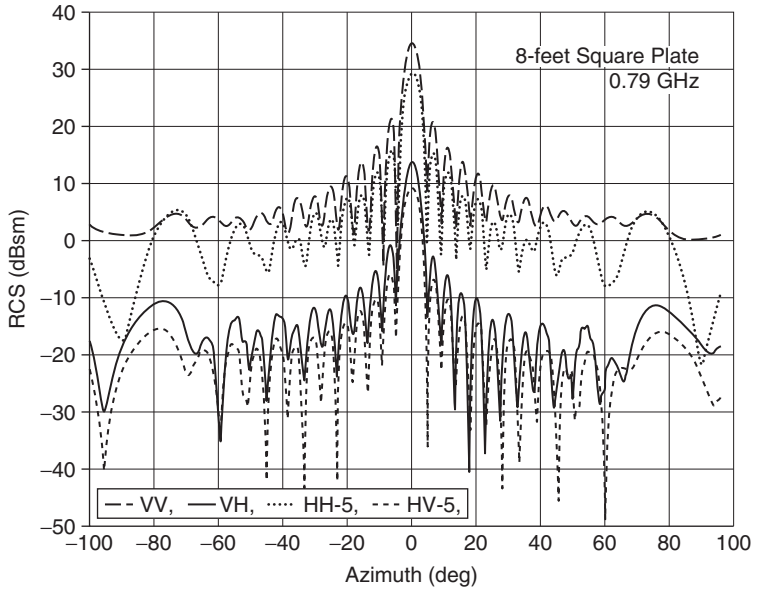


FIGURE 14.7 RCS of a square flat plate 96-in along a side measured at 790 MHz.[10] The HH and HV patterns have been artificially lowered 5 dB for clarity. (*Courtesy of S. P. Wei et al.[10], the Boeing Company. Original data courtesy of the Boeing Company, Seattle, Washington*.)

The plate is presented broadside to the incident wave at the center of the chart (0°) and is seen edge-on at the left and right sides (90°). The large specular return from the plate at the center of the chart is predicted with quite good accuracy by the flat-plate formula given in Table 14.1, later in this section. The edge-on return for VV polarization is well predicted by the straight-edge formula, also given in Table 14.1.

These undulating plate patterns follow a sin x/x variation quite closely for aspect angles out to about 30°. Beyond that angle, the two patterns differ by progressively wider margins. The sin x/x behavior is characteristic of a uniformly illuminated aperture, but unlike the one-way illumination function encountered in antenna work, the argument x for the flat plate includes a two-way (round-trip) illumination function. Thus, the beamwidth of the echo response of a flat plate is half the beamwidth of an antenna aperture of the same size. The prominent lobe in the horizontal pattern in the region between 70° and 80° is due to a surface traveling wave.

In contrast to the pattern of a flat plate, the RCS pattern of a corner reflector is quite broad. This is because the corner reflector is a reentrant structure, and no matter what its orientation (within limits, of course), internally reflected waves are directed back toward the source of the incident wave. A corner reflector is formed by two or three flat plates intersecting at right angles, and waves impinging on the first face are reflected onto the second; if there is a third face, it receives waves reflected by the first two faces. The mutual orthogonality of the faces ensures that the direction taken by waves upon final reflection is back toward the source.

The individual faces of the corner reflector may be of arbitrary shape, but the most common is an isosceles triangle for the trihedral corner; dihedral corners typically have rectangular faces. The RCS of a corner reflector seen along its axis of symmetry

TABLE 14.1 RCS Approximations for Simple Scattering Features

Scattering Feature	Orientation (1)	Approximate RCS	Notes
Corner reflector	Axis of symmetry along LOS	$4\pi A_{\text{eff}}^2/\lambda^2$	(2)
Flat plate	Surface perpendicular to LOS	$4\pi A^2/\lambda^2$	(3)
Singly curved surface	Surface perpendicular to LOS	$2\pi aL^2/\lambda$	(4)
Doubly curved surface	Surface perpendicular to LOS	$\pi a_1 a_2$	(5)
Straight plate edge	LOS perpendicular to front edge and E in plane of plate	L^2/π	(6)
Curved edge	Edge element perpendicular to LOS	$a\lambda/2$	(7)
Cone tip	Axial incidence	$\lambda^2 \sin^4(\alpha/2)$	(8)
Acute flat metal corner	LOS perpendicular to rear edge and E in plane of plate	$(\lambda/6)^2$	(9)
Acute flat metal corner	LOS along corner bisector and E in plane of plate	$(\lambda/40)^2$	(9)

NOTES:
1. LOS = line of sight
2. A_{eff} = effective area contributing to multiple internal reflections
3. A = actual area of the plate
4. a = mean radius of curvature; L = length of slanted surface
5. a_1, a_2 = principal radii of surface curvature in orthogonal planes
6. L = edge length
7. a = radius of edge contour
8. α = half angle of the cone
9. Empirical values reported by Knott, Shaeffer, and Tuley[14]

is identical to that of a flat plate whose physical area matches the effective area of the corner reflector. The magnitude of the echo may be determined by finding the polygonal areas on each face of the corner receiving waves reflected by the other faces and from which the final reflection is back toward the source. The effective area is determined by summing the projections of the areas of those polygons on the line of sight[11]; the RCS is then found by squaring that area, multiplying by 4π, and dividing by λ^2.

Figure 14.8 is a collection of RCS patterns of a trihedral corner reflector with triangular faces. The reflector was fabricated of three triangular plywood panels, metallized to enhance their surface reflectivities. The aperture exposed to the radar was, therefore, an equilateral triangle, as shown in Figure 14.9. The eight patterns in Figure 14.8 were measured with the plane of the aperture tilted above or below the line of sight by the angle ϕ.

The broad central part of these patterns is due to a triple-bounce mechanism between the three participating faces, while the "ears" at the sides of the patterns are due to the single-bounce, flat-plate scattering from the individual faces. Along the axis of symmetry of the trihedral reflector in Figure 14.9 ($\theta = 0°$, $\phi = 0°$), the RCS is $\pi L^4/3\lambda^2$, where L is the length of one of the edges of the aperture. Not shown are the echo reductions obtained when the trihedral faces are not perpendicular to each other. Those reductions depend on the size of the faces expressed in wavelengths.[12,13]

The RCS of most of the simple scattering features discussed in this section may be estimated by using the simple formulas listed in Table 14.1. The RCS of some complicated targets may be estimated by representing the target as a collection of features like those listed in Table 14.1, calculating the individual contributions and then summing the contributions coherently or noncoherently, as governed by the objective of the calculations.

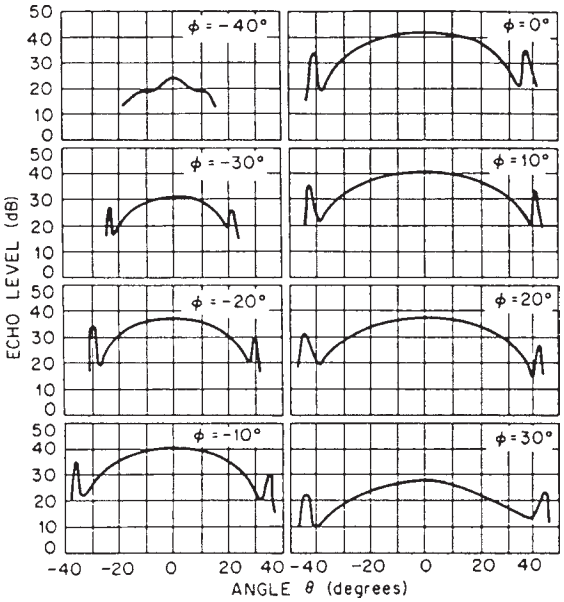


FIGURE 14.8 RCS patterns of a trihedral corner reflector. Edge of aperture = 24 in; $\lambda = 1.25$ cm. (*S. S. Robertson*[3] *© AT&T 1947. Reprinted with permission from* AT&T Technical Journal.)

Complex Objects. Objects such as insects, birds, airplanes, ships, and antennas can be much more complex than those just discussed, either because of the multiplicity of scatterers on them or because of the complexity of their surface profiles and dielectric constants. Insects are examples of the latter.

Measured values for a dozen species of insects are listed in Table 14.2. The bugs were alive for the measurements but had to be drugged to immobilize them. Figure 14.10 shows the relationship between the RCS and the mass of an insect, with the variation of a water droplet shown for comparison. Table 14.3 lists the RCS of a man as reported by Schultz et al.[16] Other comparisons have been made for both birds and insects.[17]

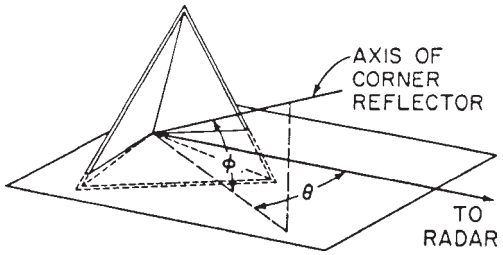


FIGURE 14.9 Coordinate system for the RCS patterns in Figure 14.8. (*S. S. Robertson*[3] *© AT&T 1947. Reprinted with permission from* AT&T Technical Journal.)

TABLE 14.2 Measured Insect RCS at 9.4 GHz[15]

Insect	Length, mm	Width, mm	Broadside RCS, dBsm	End-on RCS, dBsm
Blue-winged locust	20	4	–30	–40
Armyworm moth	14	4	–39	–49
Alfalfa caterpillar butterfly	14	1.5	–42	–57
Honeybee worker	13	6	–40	–45
California harvester ant	13	6	–54	–57
Range crane fly	13	1	–45	–57
Green bottle fly	9	3	–46	–50
Twelve-spotted cucumber beetle	8	4	–49	–53
Convergent lady beetle	5	3	–57	–60
Spider (unidentified)	5	3.5	–50	–52

NOTE: Original values reported in square centimeters have been converted here to dBsm, which is decibels relative to a square meter.

TABLE 14.3 Measured RCS of a Man[16]

Frequency, GHz	RCS, m^2
0.41	0.033–2.33
1.12	0.098–0.997
2.89	0.140–1.05
4.80	0.368–1.88
9.375	0.495–1.22

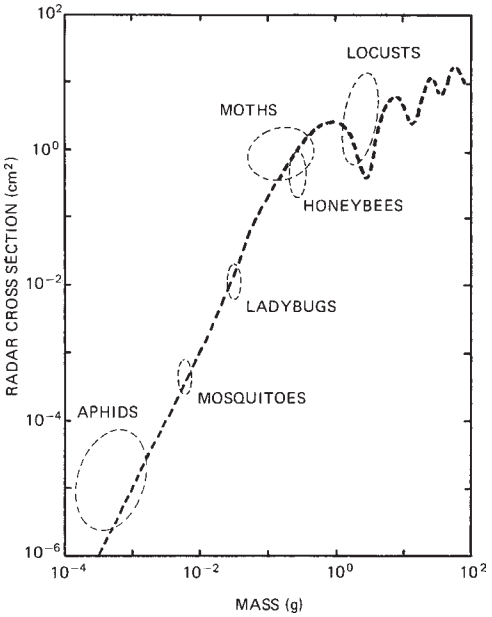


FIGURE 14.10 Sample of measured RCS of insects as a function of insect mass at 9.4 GHz, based on Riley's summary. The dashed line is the calculated RCS of water droplets for comparison. (*after J. R. Riley[18] © IEEE 1985*)

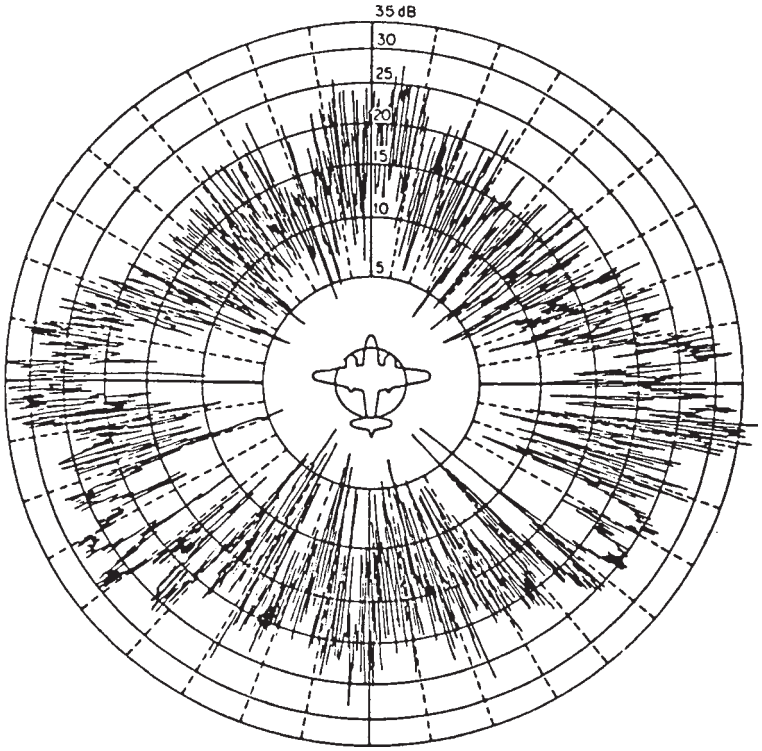


FIGURE 14.11 Measured RCS pattern of a B-26 bomber at a frequency of 3 GHz[19]

Examples of RCS patterns for aircraft are shown in Figures 14.11 and 14.12. The B-26 pattern in Figure 14.11 was measured at a frequency of 3 GHz. The polar format is useful for display purposes but is not as convenient for detailed comparisons as is a rectangular format like the one used in Figure 14.12. That rectangular pattern is of a one-third scale model C-29 aircraft and was displayed in the early 2000s on a U.S. Air Force web site.[20] The C-29 is a military version of the Raytheon Hawker 800XP mid-size business jet.

The Air Force web site reveals very little technical detail about the test conditions attending the data collection, such as the frequency and polarization of the measurements, not even the units in which the RCS data are displayed. However, even if we do not know the test frequency or polarization, we do know that the full-scale RCS will be 10 log $(3 \times 3) = 9.5$ dB higher than those charted in the figure (i.e., higher by the square of the inverse scale factor). We suspect that nose-on incidence is at the center of the pattern and that the charted RCS data are in decibels above a square meter at the test frequency.

Figure 14.13 charts the RCS of a ship measured at 2.8 and 9.225 GHz at horizontal polarization. The data were collected by a shore-based radar instrumentation complex as the ship steamed in a large circle on Chesapeake Bay. The three traces in these charts are the 80, 50, and 20 percentile levels of the signals collected over aspect angle "windows" 2° wide. The patterns are not symmetrical, especially at the higher frequency. Note that the RCS can exceed 1 mi^2 (64.1 dBsm).

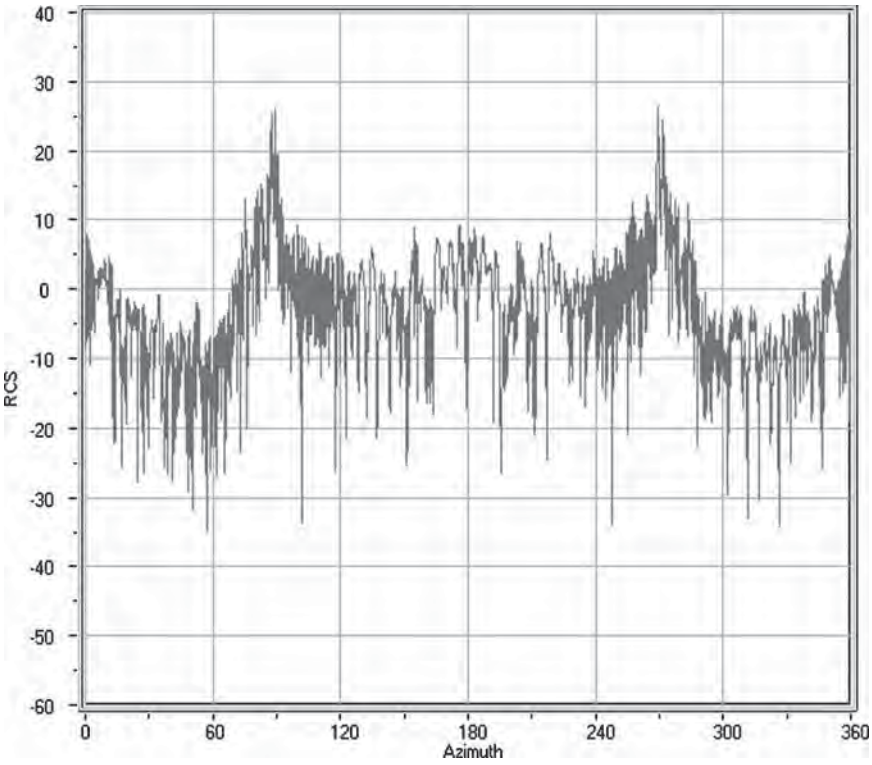


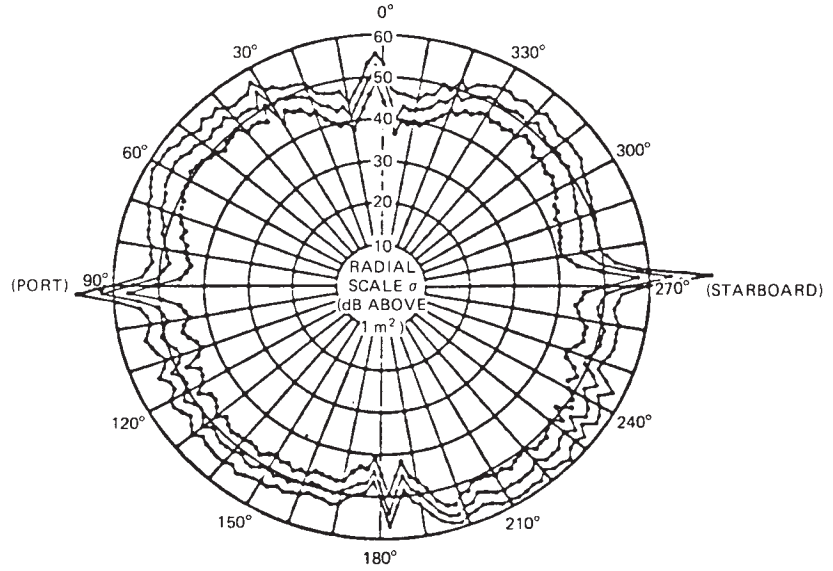
FIGURE 14.12 Measured RCS of a one-third scale model C-29 military aircraft[20]

An empirical formula for the RCS of a naval ship is

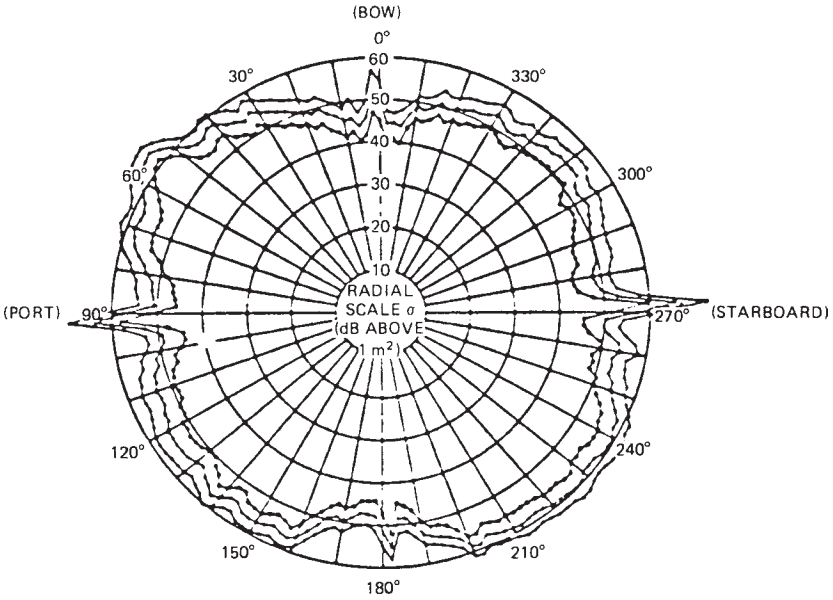
$$\sigma = 52 f^{1/2} D^{3/2} \tag{14.3}$$

where f is the radar frequency in megahertz and D is the full-load displacement of the vessel in kilotons.[21,22] The relationship is based on measurements of several ships at low grazing angles and represents the average of the median RCS in the port and starboard bow and quarter aspects, but excluding the broadside peaks. The statistics include data collected at nominal wavelengths of 3.25, 10.7, and 23 cm for ship displacements ranging from 2 to 17 kilotons.

Figure 14.14 summarizes the general RCS levels of the wide variety of targets discussed in this section, with the RCS of a metallic sphere shown as a function of its volume for comparison. The ordinate is the RCS in square meters, and the abscissa is the volume of the target in cubic feet. Because the chart is intended only to display the wide range in RCS that may be encountered in practice, the locations of targets on the chart are approximate at best. Within given target classes, the RCS may be expected to vary by as much as 20 or 30 dB, depending on frequency, aspect angle, and specific target characteristics. Readers seeking more explicit detail than this should consult referenced material at the end of this chapter.



(a)



(b)

FIGURE 14.13 Measured RCS of a large naval auxiliary ship for horizontal incident polarization. The upper pattern (*a*) is for 2.800 GHz, and the lower pattern (*b*) is for 9.225 GHz. Shown are the 80, 50, and 20 percentile levels based on the statistics of the data over 2° aspect angle windows.

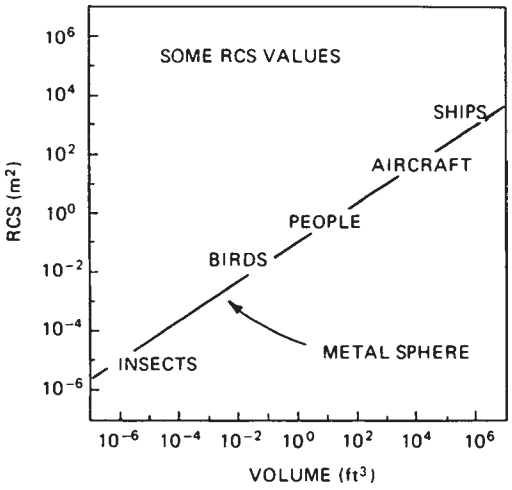


FIGURE 14.14 Summary of RCS levels of targets discussed in this section. The locations of targets on the chart are general indications only.

14.3 RCS PREDICTION TECHNIQUES

As shown in Figure 14.2, scattering obstacles are generally sorted into three different regimes based on their body size in wavelengths. Very loosely, these three regions are

- Rayleigh region: typical body size $< \lambda$
- Resonance region: $\lambda <$ typical body size $< 3\lambda$
- Optics region: $3\lambda <$ typical body size

The boundaries separating the three regimes are diffuse at best. The utility of our RCS estimation and computation methods depends in large measure on where in this size scheme we find our target.

Although the complexity and size of most scattering objects preclude the application of exact methods of radar cross-section prediction, exact solutions for simple bodies provide valuable checks for approximate methods. The exact methods are restricted to relatively simple or relatively small objects in the Rayleigh and resonant regions, whereas most of the approximate methods have been developed for the optics region, also called the *high-frequency region*. There are exceptions to these general limitations, of course. The exact solutions for many objects can be used for large bodies well into the optics region if one uses arithmetic of sufficient precision, and many optics approximations can be extended to bodies of modest electrical size into the resonance region. Low-frequency approximations developed for the Rayleigh region can extend upward into the resonance region.

Exact Methods. The exact methods are based on either the integral or differential form of Maxwell's equations.

Differential Equations. Maxwell's four differential equations constitute a succinct statement of the relationship between electric and magnetic fields produced by currents and charges and by each other.[23] The four equations may be manipulated for isotropic source-free regions to generate the wave equation:

$$\nabla^2\mathbf{F} + k^2\mathbf{F} = 0 \tag{14.4}$$

where **F** represents either the vector electric field or the vector magnetic field. Equation 14.4 is a second-order differential equation that may be solved as a boundary-value problem when the fields on the surface of the scattering obstacle are specified. The fields are typically represented as the sum of known and unknown components (incident and scattered fields), and the boundary conditions are the known relationships that must be satisfied between the fields (both electric and magnetic) just inside and just outside the surface of the obstacle. Those boundary conditions are particularly simple for solid conducting or dielectric objects.

The boundary conditions involve all three components of the vector fields, and the surface of the body must coincide with a coordinate of the geometrical system in which the body is described. For example, the coordinate r = constant could represent a spherical surface. The solution of the wave equation is most useful for those systems in which the equation is separable into ordinary differential equations in each of the variables. The scattered fields are typically expressed in terms of infinite series, the coefficients of which are to be determined in the actual solution of the problem. Once obtained, the solution allows the fields to be calculated at any point in space, which in RCS problems is the limit as the distance from the obstacle becomes infinite. The solution of the wave equation may then be used in Eq. 14.1 to determine the scattering cross section.

Aside from a few very simple objects, such as the sphere and the infinite circular cylinder, the solution for Eq. 14.4 is more academic than practical. The solution for other structures, such as infinite parabolic and elliptic cylinders, are difficult at best, and for many structures whose surfaces may coincide with a coordinate system, there is no convenient method of solution.

The most useful and practical of the exact solutions available is that of the perfectly conducting sphere, shown earlier in Figure 14.2. Metal spheres are used routinely as calibration targets for RCS measurements because an exact solution is available; metal spheres are not very hard to build; and efficient computer codes are available for obtaining the exact solution. No other scattering body affords all these conveniences.

Integral Equations. Maxwell's equations may also be manipulated to generate a pair of integral equations (known as the Stratton-Chu equations[24]):

$$\mathbf{E}_s = \oint \{ikZ_0(\mathbf{n}\times\mathbf{H})\Psi + (\mathbf{n}\times\mathbf{E})\times\nabla\Psi + (\mathbf{n}\cdot\mathbf{E})\nabla\Psi\}\, dS \tag{14.5}$$

$$\mathbf{H}_s = \oint \{-ikY_0(\mathbf{n}\times\mathbf{E})\Psi + (\mathbf{n}\times\mathbf{H})\times\nabla\Psi + (\mathbf{n}\cdot\mathbf{E})\nabla\Psi\}\, dS \tag{14.6}$$

where **n** is the unit surface normal erected at the surface patch dS and ψ is the Green's function:

$$\Psi = e^{ikr}/4\pi r \tag{14.7}$$

The distance r in Eq. 14.7 is measured from the surface patch dS to the point at which the scattered fields are desired, which could be another surface patch. These expressions state that if the total electric and magnetic field distributions are known over a completely closed surface S (as indicated by the little circle on the integral sign), the fields anywhere in space may be computed by summing (integrating) those surface field distributions over the entire surface. This scattering problem relies on the same two equations, but instead of measuring the total fields over a closed surface surrounding the body, one determines the fields induced on the body surfaces themselves by the incident wave and then solves a system of linear equations. These surface fields become unknowns to be determined. The two equations are coupled because the unknowns appear on both sides of both equations. The method of solution is known as the *method of moments (MOM)*,[25] which reduces the integral equations to a collection of homogeneous linear equations that may be solved by matrix techniques.

Once the boundary conditions have been specified, the surface S is split into a collection of discrete patches, as suggested in Figure 14.15. The patches must be small enough (typically less than $\lambda/5$) that the unknown currents and charges over each patch are constant or at least can be described by simple functions. A weighting function may be assigned to each patch, and the problem is essentially solved when the amplitude and phase of those functions have been determined.

If the point of observation is forced down to a general surface patch, the fields on the left sides of Eqs. 14.5 and 14.6 are those due to the contributions of the fields on all other patches, plus the incident fields and a "self-field." The self-field (or self-current or self-charge) is moved to the right side of the equations, leaving only the known incident field on the left side. When the process is repeated for each patch on the surface, a system of $2n$ linear homogeneous equations in $2n$ unknowns is generated. If the boundary conditions permit the decoupling of the equations, the number of unknowns may be halved (to n equations in n unknowns).

The coefficients of the resulting matrix involve only the electrical distances (in wavelengths) between all patches taken by pairs, and the orientation of patch surface normals. The unknown fields may be found by inverting the resulting matrix and multiplying the inverted matrix by a column matrix representing the incident field at each patch. The surface fields are then summed in integrals similar to Eqs. 14.5 and 14.6 to obtain the scattered field, which then may be inserted in Eq. 14.1 to compute the RCS.

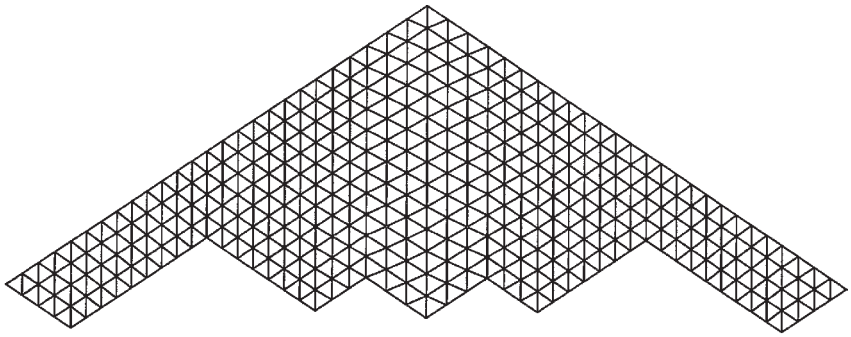


FIGURE 14.15 The method of moments divides the body surface into a collection of discrete patches. This planform of the U.S. Air Force B-2 Spirit stealth bomber uses triangular patches.

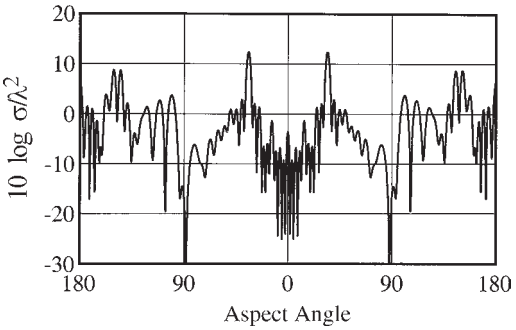


FIGURE 14.16 Azimuth-plane MOM-predicted edge-on RCS of a flat plate having the planform shown in Figure 14.15 for wing leading edges 5λ in length. The incident electric field was in the plane of the plate.

The method of moments has become a powerful tool in the prediction and analysis of electromagnetic scattering, with applications for antenna design as well as RCS prediction. This method has three limitations, however.

First, because computer memory and processing time both increase rapidly with the electrical size of the object, there might be an economic limit for the maximum target electrical size (in wavelengths) for which MOM can be used. Second, MOM yields numbers, not formulas, and is therefore a numerical experimental tool. However, trends may be established by running these numerical experiments repeatedly for parametric changes in the geometry or configuration of an object, or in the angle of arrival or of the frequency, of the incident wave. Third, the solutions for some objects may contain spurious resonances that do not actually exist, thereby reducing the confidence one may have in applying the method to arbitrary structures.

Figure 14.16 shows a MOM-code prediction of the edge-on RCS pattern of a large, flat metal plate of zero thickness having the planform of Figure 14.15. For the purpose of illustration, we chose a simulation frequency such that the leading edges of the wings were 5λ long. The incident polarization and the direction of the incident wave were both in the plane of the plate. Nose-on incidence lies at zero-degree aspect at the center of the chart, and the tail-on aspect lies at 180° at both sides of the chart.

As shown in Table 14.1, the approximate RCS of a straight edge of length L presented perpendicular to the incident wave is $\sigma = L^2/\pi$. However, this estimate is some 3 dB lower than the peak amplitude attained in Figure 14.16 by the leading-edge echoes 34° either side of nose-on incidence. Evidently there are other, more subtle echo sources that contribute to the RCS at these aspects, possibly surface traveling-wave contributions.

Approximate Methods. Approximate methods for computing scattered fields are available in both the Rayleigh and the optics regions. Rayleigh-region approximations may be derived by expanding the wave equation (Eq. 14.4) in a power series of the wavenumber k.[26] Higher-order terms of the expansion become progressively more difficult to obtain. The RCS pattern of a Rayleigh scatterer is very broad, especially if the object has similar transverse and longitudinal dimensions. The magnitude of the echo is proportional to the square of the volume of the object and varies as the fourth

power of the frequency of the incident wave.[27] Because the method of moments is quite well suited to the solution of Rayleigh-region problems, analytical Rayleigh-region expansions for predicting the RCS of electrically small objects have lapsed into disuse.

Several approximate methods have been devised for the optics region, each with its particular advantages and limitations. The most mature of the methods are *geometric optics* and *physical optics*, with later methods attacking the problem of diffraction from edges and from shadow boundary field discontinuities. While the general accuracy of the optics region approximations improves as the scattering obstacle becomes electrically larger, some of them give reasonably accurate results (within 1 or 2 dB) for objects as small as a wavelength or so.

Geometric Optics. The theory of geometric optics (GO) is based on the conservation of energy within a slender fictitious tube called a *ray*. The direction of propagation is along the tube, and contours of equal phase are perpendicular to it. In a lossless medium, all the energy entering the tube at one end must come out the other end, but energy losses within the medium may also be accounted for. An incident wave may be represented as a collection of a large number of rays, and when a ray strikes a surface, part of the energy is reflected and part is transmitted across the surface. The amplitude and phase of the reflected and transmitted rays depend on the properties of the media on either side of the surface. The reflection is perfect if the surface is perfectly conducting, and no energy is transmitted across the boundary into the body. When energy can pass through the surface, transmitted rays are bent toward the surface normal in crossing a surface into an electrically denser medium (higher index of refraction) and away from the surface normal into a less dense medium. This bending of rays is known as *refraction*.

Depending on surface curvature and body material, reflected and transmitted rays may diverge from one another or they may converge toward each other. This dependence is the basis for the design of lenses and reflectors at radar wavelengths as well as at optical wavelengths. The reduction in intensity as the rays diverge (spread away) from the point of reflection can be calculated from the curvatures of the reflecting surface and the incident wave at the *specular point*, which is that point on the surface where the angle of reflection equals the angle of incidence. The principal radii of curvature of the surface are measured in two orthogonal planes at the specular point, as shown in Figure 14.17. When the incident wave is planar and the direction of interest is back toward the source, the geometric optics RCS is simply

$$\sigma = \pi a_1 a_2 \tag{14.8}$$

where a_1 and a_2 are the radii of curvature of the body surface at the specular point.

This formula becomes exact in the optical limit of vanishing wavelengths and is probably accurate to 10 or 15 percent for radii of curvature as small as 2λ or 3λ. It assumes that the specular point is not close to an edge. When applied to dielectric objects, the expression should be multiplied by the square of the voltage reflection coefficient associated with the material properties of the object. Internal reflections may also be accounted for, and the phase of internally reflected rays should be adjusted according to the electrical path lengths traversed within the body material. The net RCS should then be computed as the coherent sum of the surface reflection plus all significant internal reflections. Equation 14.8 fails when one or both surface radii of

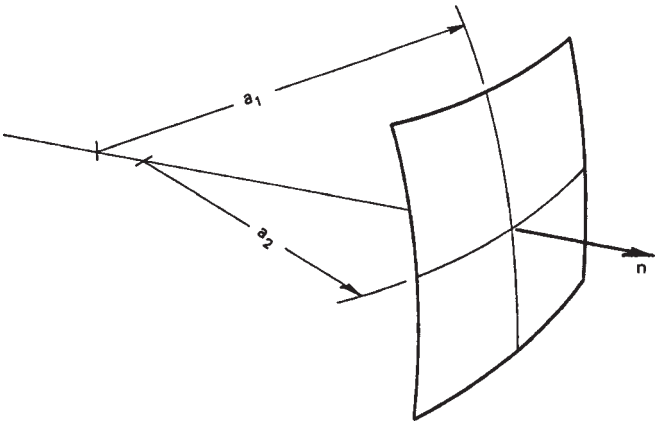


FIGURE 14.17 The geometric optics RCS of a doubly curved surface depends on the principal radii of curvature at the specular point. The specular point is where the surface normal points toward the radar.

curvature at the specular point become infinite, yielding infinite RCS, which is obviously wrong. This occurs for flat and singly curved surfaces.

Physical Optics. The theory of physical optics (PO) is a suitable alternative for bodies with flat and singly curved surface features. The theory is based on two approximations in the application of Eqs. 14.5 and 14.6, both of which are reasonably effective in a host of practical cases. The first is the *far-field approximation*, which assumes that the distance from the scattering obstacle to the point of observation is large compared with any dimension of the obstacle itself. This allows us to replace the gradient of Green's function with

$$\nabla\psi = ik\psi_o \mathbf{s} \tag{14.9}$$

$$\psi_o = e^{-ik\mathbf{r}\times\mathbf{s}}\, e^{ikR_o}/4\pi R_o \tag{14.10}$$

where $\mathbf{r}$ is the position vector of integration patch dS and $\mathbf{s}$ is a unit vector pointing from an origin in or near the object to the far-field observation point, usually back toward the radar. R_o is the distance from the origin of the object to the far-field observation point.

The second is the *tangent plane approximation*, in which the tangential field components $\mathbf{n}\times\mathbf{E}$ and $\mathbf{n}\times\mathbf{H}$ are approximated by their geometric optics values. That is, a tangent plane is passed through the surface coordinate at the patch dS, and the total surface fields are taken to be precisely those that would have existed had the surface at dS been infinite and perfectly flat. In essence, we do not know these fields, but we take our best guess as to what they are and insert that estimate into either of the two integrals. That done, the estimate of the unknown fields in the integrals of Eqs. 14.5 and 14.6 may be expressed entirely in terms of the known incident field values. The problem then becomes one of evaluating the chosen integral and substituting the result into Eq. 14.1 to obtain the RCS.

If the surface is a good conductor, the total tangential electric field is virtually zero and the total tangential magnetic field is twice the amplitude of the incident tangential magnetic field:

$$\mathbf{n} \times \mathbf{E} = 0 \tag{14.11}$$

$$n \times \mathbf{H} = \begin{cases} 2n \times \mathbf{H}_i & \text{illuminated surfaces} \\ n & \text{shaded surfaces} \end{cases} \tag{14.12}$$

Note that the tangential components of both the electric and the magnetic fields are set to zero over those parts of the surface shaded from the incident field by other body surfaces. Other approximations may be devised for nonconducting surfaces. If the incident wavelength is long enough, for example, the surface of a soap bubble or the leaf of a tree may be modeled as a thin membrane, over which neither the electric nor magnetic fields are zero.

The physical-optics integral is easy to evaluate for flat metallic plates because the phase is the only quantity within the integral that varies, and it varies linearly across the surface. When applied to a rectangular metal plate, the integral evaluation leads to the RCS:

$$\sigma = 4\pi \left| \frac{A \cos\theta}{\lambda} \cdot \frac{\sin(k\ell \sin\theta \cos\phi)}{k\ell \sin\theta \cos\phi} \cdot \frac{\sin(kw \sin\theta \sin\phi)}{kw \sin\theta \sin\phi} \right|^2 \tag{14.13}$$

where $A = \ell w$ is the physical area of the plate, θ is the angle between the surface normal of the plate and the direction to the radar, ϕ is the angle between the plane containing the line of sight and the edge whose length is ℓ, and w is the width of the plate. (A more general physical optics formula is available for the bistatic scattering from any polygonal plate.[28,29])

If we set $\phi = 0°$ or 90°, we obtain a *principal plane* RCS pattern (incidence in a plane perpendicular to a pair of edges). When $\phi = 0°$, Eq. 14.13 becomes

$$\sigma = 4\pi \left| \frac{A \cos\theta}{\lambda} \cdot \frac{\sin(k\ell \sin\theta)}{k\ell \sin\theta} \right|^2 \tag{14.14}$$

If we set $\phi = 90°$ instead of $\phi = 0°$, we get nearly the same answer, except that $k\ell$ in Eq. 14.14 becomes kw. The physical optics integral is not dependent on the polarization of the incident wave and is unreliable for angles much greater than 30° from broadside incidence.

By way of comparison, the physical optics formula for the RCS of a circular metal disk is

$$\sigma = 16\pi \left| \frac{A \cos\theta}{\lambda} \cdot \frac{J_1(kd \sin\theta)}{kd \sin\theta} \right|^2 \tag{14.15}$$

where A is the physical area of the disk, d is its diameter, and $J_1(x)$ is the Bessel function of the first kind of order 1. Equations 14.13 through 14.15 all reduce to the value listed in Table 14.1 for normal incidence. For further comparison, the PO patterns of two square plates and a circular disk are plotted in Figure 14.18.

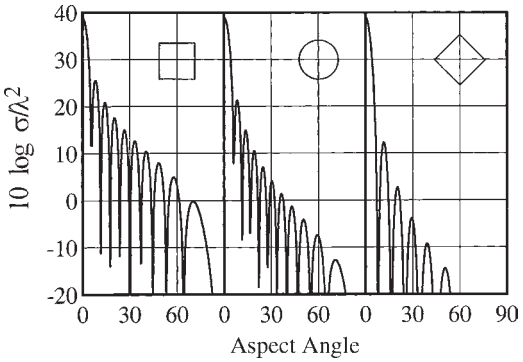


FIGURE 14.18 PO patterns of the RCS of a square plate, a disk, and second square plate

The three patterns each cover a 90° sector from broadside to edge-on and are placed side-by-side for clarity. The areas of all three plates were fixed at $25\lambda^2$, hence all three patterns rise to the same amplitude at broadside incidence (zero aspect). The center pattern is for the disk, while the first and third are both for a square plate. However, the square plate was oriented for a principal-plane pattern in the leftmost chart and like a diamond ($\phi = 45°$) in the rightmost chart. The amount of surface distributed toward the sides of the plates influences the sidelobe levels.

The PO integral is somewhat more complicated to evaluate when the surface is singly or doubly curved. An exact evaluation can be performed for a circular cylinder and a spherical cap viewed along the axis of symmetry, but not for a truncated cone or a spherical cap seen along other than the axis of symmetry. Even so, the exact evaluation for the cylinder includes fictitious contributions from the shadow boundaries at the sides of the cylinder that do not appear in a *stationary phase approximation*.[30]

The amplitude of the elemental surface contributions changes slowly over the surface of integration whereas the phase changes much more rapidly. As such, the net contribution in regions of rapid phase change is essentially zero and may be ignored. As the specular regions are approached, on the other hand, the phase variation slows down and then reverses as the specular point is crossed. This results in a nonzero specular contribution to the integral. The phase variation near the shadow boundaries is rapid, hence, surface contributions there are ignored in a stationary phase evaluation, but an exact evaluation includes them because the shadow boundaries are the limits of integration. Because the actual surface field distributions do not suddenly drop to zero as the shadow boundary is crossed, as assumed by the theory, the shadow boundary contributions are spurious.[31,32] Thus, a stationary phase approximation of the physical optics integral over closed curved surfaces tends to be more reliable than an exact evaluation of the integral.

With this in mind, the stationary phase result for a circular cylinder is

$$\sigma = ka\ell^2 \left| \frac{\sin(k\ell \sin\theta)}{k\ell \sin\theta} \right|^2 \tag{14.16}$$

where a is the radius of the cylinder, ℓ is its length, and θ is the angle off broadside incidence. Equation 14.16 includes only the contribution from the curved side of the

cylinder and not its flat ends, which may be included by using the prescription of Eq. 14.15. Equation 14.16 may also be used to estimate the RCS of a truncated right circular cone if the radius a is replaced by the mean radius of the cone and ℓ is replaced by the length of the slanted surface.

Although the theory of physical optics offers a significant improvement over geometric optics for flat and singly curved surfaces, it suffers other drawbacks. Even though we obtain the proper result for most of the illuminated surface, the physical optics integral yields false contributions from the shadow boundaries, as already noted. Moreover, physical optics shows no dependence on the polarization of the incident wave and yields different results when the receiver and the transmitter are interchanged. These effects contradict observed behavior. Finally, it errs by wider margins as the direction of observation moves farther away from the specular direction. Keller's *geometrical theory of diffraction* (GTD) offers an improvement in both the polarization dependence and the predicted values in the wide-angle regions.[33,34]

Geometric Theory of Diffraction. GTD is a ray-tracing method that assigns an amplitude and phase to fields diffracted at smooth shadow boundaries and at surface discontinuities. Because the latter are much more significant in backscattering computations than the former, we focus here on edge diffraction. The theory assumes that a ray striking an edge excites a cone of diffracted rays, as in Figure 14.19. The half angle of this *diffraction cone* is equal to the angle between the incident ray and the edge. Unless the point of observation lies on the diffraction cone, no value is assigned the diffracted field. The scattering direction in backscattering problems is the reverse of the direction of incidence, whence the diffraction cone becomes a disk, and the scattering edge element is perpendicular to the line of sight.

The amplitude of the diffracted field is given by the product of a *diffraction coefficient* and a *divergence factor*, and the phase depends on the phase of the edge excitation and on the distance between the observation point and the diffracting edge element. Two cases are recognized, depending on whether the incident field is polarized parallel or perpendicular to the edge.

The diffracted field is given by the formula:

$$E_d = \frac{\Gamma e^{iks} e^{i\pi/4}}{\sqrt{2\pi ks}\,\sin\beta}(X \mp Y) \tag{14.17}$$

where Γ is a divergence factor, X and Y are diffraction coefficients, β is the angle between the incident ray and the edge, and s is the distance to the observation point from the point of diffraction. The difference of the two diffraction coefficients is

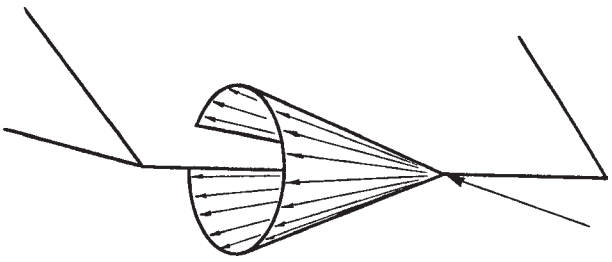


FIGURE 14.19 The Keller cone of diffracted rays

used when the incident electric field is parallel to the edge (*TM polarization*) and the sum when the incident magnetic field is parallel to the edge (*TE polarization*).

The divergence factor accounts for the decay in amplitude as the rays spread away from the edge element and includes the effects of the radius of the edge if it is curved, as at the end of a truncated cylinder, and the radius of curvature of the incident phase front.[35] The divergence factor for a two-dimensional edge (of infinite length) illuminated by a plane wave is $\Gamma = 1/s$. The diffraction coefficients are

$$X = \frac{\sin(\pi/n)/n}{\cos(\pi/n) - \cos[(\phi_i - \phi_s)/n]} \tag{14.18}$$

$$Y = \frac{\sin(\pi/n)/n}{\cos(\pi/n) - \cos[(\phi_i + \phi_s)/n]} \tag{14.19}$$

where ϕ_i and ϕ_s are the angles of the planes of incidence and scattering, as measured from one face of the wedge (say, the illuminated one), and n is the exterior wedge angle normalized with respect to π. When these expressions are evaluated for the case of a flat metal edge viewed edge-on with the incident polarization in the plane of the plate, they yield the edge-on formula $\sigma = L^2/\pi$ listed in Table 14.1.

Keller's diffraction coefficients are based on an approximation of the exact solution for an infinite (two-dimensional) metal wedge as applied to a three-dimensional problem.[36] Although this adaptation of a two-dimensional solution to the three-dimensional world is reasonably effective most of the time, the diffraction coefficients inconveniently blow up just when needed most. A cursory examination of Eqs. 14.18 and 14.19 will show why.

The denominators of both expressions contain the difference of two cosine terms that become equal in two different cases. When the scattering direction ϕ_s is aligned along the shadow boundary where $\phi_i - \phi_s = \pi$, the diffraction coefficient X in Eq. 14.18 becomes singular, a meaningless result. When the scattering direction ϕ_s is aligned along the specular direction where the local angle of reflection is equal to the local angle of incidence, then $\phi_i + \phi_s = \pi$. In this case, it is the diffraction coefficient Y in Eq. 14.19 that becomes singular, a similarly meaningless result. Note that these two singularities do not depend on body geometry, but only on the relative dispositions of the incidence and scattering directions.

Physical Theory of Diffraction. The singularities in GTD are overcome in the *physical theory of diffraction* (PTD) formulated by P. Ia. Ufimtsev.[37,38] (Although these publications may be difficult to find, we cite them here for completeness. *Note from the editor*: The reader interested in this subject might also see the Ufimtsev paper "Comments on Diffraction Principles and Limitations of RCS Reduction Techniques," *Proc. IEEE*, vol. 84, pp. 1830–1851, December 1996.) Like Keller, Ufimtsev relied on the approximate (wide-angle) solution of the two-dimensional wedge problem, but he distinguished between "uniform" and "nonuniform" induced surface currents. The uniform currents were none other than the surface currents of physical optics, whereas the nonuniform currents were taken to be undefined filamentary currents along the edge itself. Ufimtsev never attempted to work out his filamentary *fringe* currents, but instead traced their effect directly to the far scattered field.

Recognizing that the far-field PO contribution to the far field was the part of the GTD prescription giving rise to singularities in the X and Y of Eqs. 14.18 and 14.19, Ufimtsev devised a modified set of diffraction coefficients by simply subtracting away

the offensive PO diffraction coefficients from the time-honored wide-angle wedge solution. This generated a new set of diffraction coefficients that retained only the edge terms, therefore excluding any surface terms. Ufimtsev's PTD coefficients were well behaved in almost all directions in space, but suffered one disadvantage: in order to calculate the RCS of an arbitrary edged body, one had to sum all the PTD edge contributions, *plus* all the PO and GO surface contributions. However, the procedure is viable and has been well documented.[40]

Incremental Length Diffraction Coefficient. GTD and PTD are both based on the exact solution of the two-dimensional wedge problem, for which the directions of incidence and scattering are perpendicular to the edge. When extended to the case of oblique incidence, the direction of observation must lie along a generator of the Keller cone depicted in Figure 14.19. If the edge is straight and of finite length, as in the three-dimensional world, Eq. 16 of Knott[1] provides an approximation of the RCS. If the edge is curved, it may be regarded as a collection of infinitesimally short segments butted together, and the scattered fields may be computed via an integration of incremental fields diffracted by each element of the edge. This is the concept introduced by Mitzner,[41] and the summation of the fields diffracted by the edge elements implies an integral around the edge contour.

Mitzner, however, sought the fields scattered in arbitrary directions, not just those along the local Keller cones, and for this purpose he developed his *incremental length diffraction coefficient* (ILDC). Extending the example provided by Ufimtsev, he devised a set of diffraction coefficients for arbitrary directions of incidence and scattering. Not unexpectedly, those coefficients are more complicated than the *Xs* and *Ys* appearing in Eqs. 14.18 and 14.19.

Mitzner expressed his result as the diffracted electric-field components parallel and perpendicular to the plane of scattering in terms of the components of the incident electric field parallel and perpendicular to the plane of incidence. As such, the diffraction coefficients may be expressed as three separate pairs representing parallel-parallel, perpendicular-perpendicular, and parallel-perpendicular (or perpendicular-parallel) combinations. One member of each pair is due to the total surface current on the diffracting edge (including the assumed filamentary edge currents), and the other is due to the uniform physical optics currents. Mitzner subtracted one member of each pair from the other, thereby retaining the contributions from the filamentary edge currents alone.

The results have the identical form of Ufimtsev's expressions, in which the PO coefficients are subtracted from the non-PO coefficients. Thus, Mitzner's expression for the scattered field contains only the contributions from the filamentary edge currents. In applying his theory to scattering objects, therefore, the contributions of nonfilamentary-induced surface currents must be accounted for separately, just as in Ufimtsev's PTD. When the directions of incidence and scattering become perpendicular to an edge, the perpendicular-parallel terms disappear and Mitzner's diffraction coefficients then reduce identically to Ufimtsev's.

Method of Equivalent Currents. Undertaking what he called a more rigorous evaluation of the fields induced on a wedge, Michaeli duplicated Mitzner's result for the total surface currents, confirming Mitzner's prior development, but he did not explicitly remove the PO surface-current contributions.[42,43] Thus, like Keller's *X* and *Y*, Michaeli's diffraction coefficients become singular in the transition regions along

the reflection and shadow boundary directions. Michaeli later investigated the removal of the singularities, the cleverest of which was the use of a skewed coordinate system along the wedge surfaces.[44,45]

While these methods of evaluating the fields scattered by edge elements may be applicable to smooth unbounded edges, they do not account for the discontinuities at corners where the edges turn abruptly in other directions. An attack on the corner problem has been suggested by Sikta et al.[46]

14.4 RCS MEASUREMENT TECHNIQUES

RCS measurements may be required for any of several reasons, ranging from scientific inquiry to verification of compliance with product specifications. There are no formal standards governing instrumentation and measurement methods, but informal standards of good measurement practice have been recognized for decades. Depending on the size of the test object, the frequencies to be used, and other test requirements, measurements may be made in indoor test facilities or on outdoor ranges. Because one is seldom interested in the RCS of an object for only one aspect angle, all static test ranges use turntables or rotators to vary the target aspect angle. Although the purpose of testing often governs how the measurements will be made, Mack and Dybdal provide good overall guides for routine RCS testing.[47,48]

General Requirements. The most important requirement for RCS measurements is that the test object be illuminated by a radar wave of acceptably uniform amplitude and phase. Good practice dictates that the amplitude of the incident wave deviate by no more than 0.5 dB over the transverse and longitudinal extent of the target and that the phase deviation be less than 22.5°. It is standard practice at some test ranges to physically probe the incident field at the onset of a test program to verify the amplitude uniformity of the incident wave.

The phase requirement is the basis of the far-field range criterion:

$$R > 2D^2/\lambda \tag{14.20}$$

where R is the distance between the instrumentation radar and the test object and D is the maximum target dimension transverse to the line of sight. All other error sources being fixed, compliance with the far-field requirement is generally felt to yield data with an accuracy of 1 dB or better.[49] Figure 14.20 illustrates the far-field requirement for a variety of frequencies and target sizes.

Errors attributable to radar instrumentation should be held to 0.5 dB or less, which requires careful design and selection of components. The drift in system sensitivity should not exceed this value for the time it takes to record a single RCS pattern, which sometimes may approach an hour. The dynamic range of the system should be at least 40 dB, with 60 dB preferred. Linearity over this range should be 0.5 dB or better, and if not, steps should be taken to correct measured data via calibration of the receiver transfer function (gain characteristics).

RCS measurements should be calibrated by the *substitution method*, in which an object of known scattering characteristics is substituted for the target under test. Given the known (measured or calibrated) receiver gain characteristics, this establishes the constant by which a receiver output indication may be converted to an absolute RCS value.

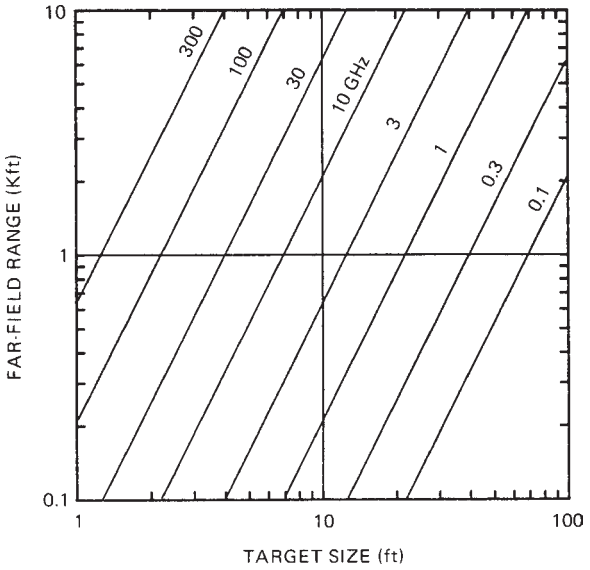


FIGURE 14.20 The far-field distance (*Reprinted with permission of SciTech Publishing, Inc.*[50])

Common calibration targets include metal spheres, right circular cylinders, flat plates, and corner reflectors. The radar cross sections of these objects may be calculated by using the expressions given in Section 14.3.

Because residual background reflections contaminate the desired target echo signal, they should be minimized by careful range design and operation. Interior walls in indoor test chambers must be covered with high-quality radar-absorbing material, and the surface of the ground on outdoor ranges should be smooth and free of vegetation. Target support structures should be designed specifically for low echo characteristics.

The effects of undesired background signals are illustrated in Figure 14.21. Because the relative phase between the background signal and the target signal is unknown, two curves are shown; they correspond to perfect in-phase and out-of-phase conditions. If the background signal is equal to the target signal (ratio of 0 dB) and the two are in phase, the total received power is four times the power due to either one. This is the value shown at the upper left of the chart (6 dB). If the two are out of phase, they cancel each other, and there is no signal at all (off the lower left of

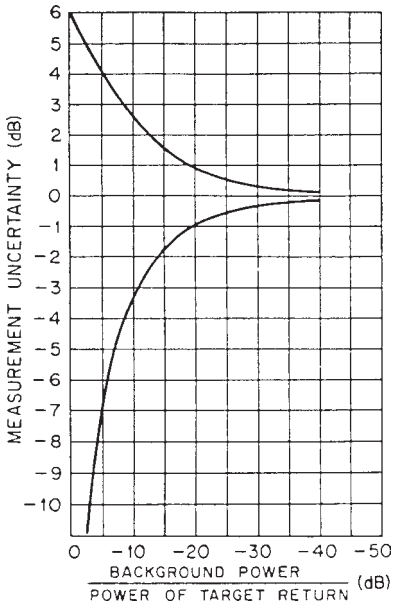


FIGURE 14.21 Measurement error as a function of the relative background power level

the chart). The chart shows that if the error due to background signals is to be 1 dB or less, the background must be at least 20 dB below the signal being measured.

Three different kinds of support structures have been demonstrated to be useful in RCS measurements. They are the low-density plastic foam column, the string suspension harness, and the slender metal pylon. The echo from a plastic foam column arises from two mechanisms. One is a coherent surface reflection, and the other is a noncoherent volume contribution from the thousands of internal cells comprising the foam material.[51,52] The column should be designed so that its surfaces are never closer than 5° to 10° to the line of sight to the radar (depending on frequency), thereby minimizing the effect of the surface reflection. The noncoherent volume return is irreducible, however, and is not influenced by the orientation of the column. The volume return of suitable foam column support materials is of the order of 1.6×10^{-6} m^2 per ft^3 of exposed column at 10 GHz.[53]

String suspension methods are best implemented indoors, where an overhead support point is normally available, although one documented design was seriously considered for outdoor use.[54] One of three configurations may be selected, all requiring a custom-made sling or harness to support the target. The first uses a single overhead support point and guy lines to a floor-mounted turntable to rotate the target. The second configuration suspends the target from an overhead turntable, reducing the guy lines and string loads at the expense of a more costly installation. The third configuration is the most costly, using a pair of turntables slaved together, one in the ceiling and one on the floor.

The echo signal from a string depends on the length and diameter of the string, its tilt angle with respect to the incident wave, and its dielectric constant. No matter what the tilt of the string, it will be presented normal to the line of sight twice in a complete rotation of the target and may cause a spike in the RCS pattern that could be erroneously attributed to the target unless otherwise accounted for. The RCS of a string rises with the fourth power of its diameter in the Rayleigh region (see Figure 14.4), and for a given tensile strength, the diameter rises only as the square root of the load to be supported. Thus, because the echo signal increases with the square of the load-carrying capacity, string suspension techniques are best suited for measurements of light objects at low frequencies.

The metal target support pylon was first suggested in 1964,[54] but a practical implementation of the concept did not appear until 1976. The configuration of the pylon is shown in Figure 14.22, and it owes its electromagnetic performance to the sharpness of its leading edge and its tilt toward the radar (to the left in the diagram). Pylons as tall as 95 ft have been built, and it is customary to treat them with radar-absorbing material to suppress the echoes from the leading and trailing edges.

The obvious advantage of the metal pylon is its superior weight-carrying capability compared with that of strings and plastic foam columns. However, because the top of the pylon is small, the rotation mechanism needed to vary the aspect angle of the target must be embedded in the target itself. This usually destroys the operational value of the target. Most of the rotators for these pylons are dual-axis, azimuth-over-elevation designs. When measurements are made with the azimuth rotation angle tilted back (away from the radar), parts of the target may sweep through the shadow cast by the top of the pylon, possibly degrading the measurements. One way to avoid this is to invert the target and tilt the rotation axis toward the radar instead of away from it. This requires the installation of the rotator in the top of the target as well as in the bottom. The unused internal cavities created for such installations must be concealed by covers or shields.

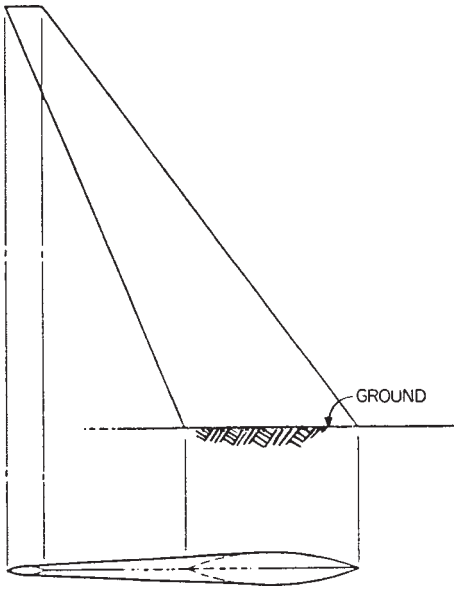


FIGURE 14.22 The metal support pylon. The design is for an incident wave arriving from the left. (*Reprinted with permission of SciTech Publishing, Inc.*[55])

It is often necessary to measure scale models, which requires the application of scaling laws. Because nonconducting materials must be scaled differently than good conductors, it is not possible to satisfy all the scaling requirements for arbitrary targets composed of conducting and nonconducting materials. Most targets requiring scale-model testing, however, are dominantly metallic, for which the perfectly conducting scaling law is generally regarded as adequate.

When normalized with respect to the square of the wavelength, the RCS patterns of two perfectly conducting objects of identical shape but different size will be identical if the objects are the same number of wavelengths in size. If a model is one-tenth of full scale, for example, it should be measured at one-tenth of the full-scale wavelength (ten times the full-scale frequency). The RCS of the full-scale target may be obtained from the scale-model measurements by multiplying the scale-model RCS by the square of the ratio of the two frequencies. In this example, that factor is 10^2, or 20 dB.

Outdoor Test Ranges. Outdoor test ranges are required when test targets are too large to be measured indoors. The far-field criterion often requires that the range to the target be several thousand feet (see Figure 14.20). Because the typical target height above the ground is a few dozen feet at best, the elevation angle to the target as seen from the radar is 1° at most and often less. At such low grazing angles, the ground is strongly illuminated by the antennas, and unless the ground bounce can be suppressed, the target will be illuminated by a multipath field. In the design of an outdoor test range, therefore, a decision must be made whether to exploit the ground bounce or to attempt to defeat it. It is generally easier to exploit it than to eliminate it.

Test ranges designed to exploit the multipath effect may be asphalted to improve the ground reflection, although many ranges are operated over natural soil. Paving the range ensures uniformity in the characteristics of the ground plane from day to day and extends its operational usefulness to higher frequencies than might otherwise be possible. A conducting screen embedded in the asphalt may improve the reflection. Paving also reduces maintenance of the ground plane, such as might be required by periodically removing vegetation and smoothing out windblown ridges in unstable soil.

The angle of incidence and the dielectric properties of asphalt and natural soil are such that the phase of the voltage reflection coefficient is within a few degrees of 180°. This being the case, one can usually choose a combination of target and antenna heights such that the wave reflected by the ground arrives at the target in phase with the wave propagated directly from the antennas. The result is the following rule for selecting the antenna and target heights:

$$h_a h_t = \lambda R/4 \tag{14.21}$$

where h_a and h_t are the antenna and targets heights, respectively, and R is the range to the target.

Because most test ranges have turntables or target pylons installed at a few fixed locations relative to a permanent radar complex, the range R is usually restricted to a few preset values. The target is installed at a height h_t high enough to minimize spurious interactions with the ground, yet low enough to minimize the size and complexity of the target support structure. Therefore, it is the antenna height h_a that is most readily controlled and adjusted to optimize the location of the first lobe in the vertical multipath interference pattern. This is easily accomplished by mounting the radar antennas on carriages that can be raised or lowered along rails installed on the side of a building or a tower.

The ideal ground plane offers a theoretical sensitivity enhancement of 12 dB over identical measurements made in free space. The actual enhancement is usually less than this, primarily because of the directivity of the antennas and imperfections in the ground plane. Antenna directivity precludes the target ever being squarely aligned on the boresights of both the real antenna and its image in the ground plane at the same time, and the reflection coefficient of typical ground planes varies from 95% to as low as 50% or less. For all except very high and very low frequencies (millimeter wavelengths and VHF), typical sensitivities are of the order of 7 to 10 dB above free space instead of the ideal 12 dB.

When the range to the target is relatively short and tests must be performed over a wide range of frequencies, it is sometimes advantageous to attempt to defeat the ground-plane effect. One option is to install a berm shaped like an inverted *V* running between the radar and the target. The purpose of the berm's slanted top is to deflect the ground-reflected wave out of the target zone. Another option is to install a series of low *radar fences* across the range. The design objective is to block ground-reflected rays from reaching the target from the radar, and vice versa, by shielding the specular zone on the ground from both. The near sides of the fences should be slanted to deflect energy upward (out of the target test zone) and may be covered with absorbing material. It is difficult, however, to prevent diffraction of radar energy from the top of the fences from reaching the target or to prevent target-diffracted signals from reaching the radar receiver via the same kind of mechanism.

Because of the large distances from the radar to the target on outdoor ranges, instrumentation radars at one time developed peak signal powers ranging from 1 to 100 kW.

The high-power instrumentation of yesteryear has been largely replaced on static test ranges with coherent stepped-frequency radars that are far more versatile, but high-power systems remain useful for dynamic testing of targets flying or sailing past them on dynamic test ranges. Stepped- and swept-frequency instrumentation systems on static ranges can collect RCS patterns at hundreds of frequencies for a single rotation of the test target. The signal-to-noise ratio can be improved when necessary by means of multisweep or multistep signal integration schemes. The price paid for much of this versatility is the requirement for more active test time per target rotation, thereby increasing measurement costs.

Indoor Test Ranges. Indoor test ranges offer protection from weather and therefore more productive testing, but unless a very large facility is available, maximum target sizes are limited to one or two dozen feet. Because of their proximity, the walls, floor, and ceiling must be covered with high-quality absorbing material. The lower the intended frequency of operation, the more expensive the absorber becomes. Absorber reflectivity ratings of −50 dB are common among the materials used. This performance is usually achievable only with the pyramidal design.[56]

Early indoor chambers were rectangular in shape, and despite the installation of good absorbent materials on the walls, RCS measurements could be contaminated by wall reflections. The most sensitive part of the anechoic chamber is the rear wall, which receives 95 to 99 percent of the power radiated by the radar; hence, the best absorber should be reserved for the rear wall.[57] The floor, ceiling, and sidewalls also contribute errors, via a quadruplet of reflections not unlike those due to the ground plane of outdoor ranges. A remedy is the tapered anechoic chamber, which eliminates most of the sidewall reflections purely by tilting portions of the walls, floor, and ceiling away from the chamber centerline.[58,59,60]

Even targets of modest size cannot be measured at the far-field distance in indoor facilities because most chambers are not much more than 200 ft or so in length. It is possible, however, to provide the necessary uniformity of illumination by *collimating* the radiated beam. This can be done by inserting a lens between the radar and the target[61,62] or by reflecting the radar beam off a collimating reflector. This latter concept is known as the *compact range* because a beam of parallel rays can be generated in a much shorter distance than would be possible without the collimating device.

The reflector offers a different way to collimate a beam. In contrast to the lens, which is placed between the radar and the test object, the radar and the test object remain on the same side of the reflector, as shown in Figure 14.23. The reflector is typically an offset paraboloid, meaning the paraboloidal surface does not include the vertex of the generating parabola. This permits the feed that excites the reflector to be placed out of the beam reflected toward the target. If the test object is held within one or two focal lengths of the reflector and if the reflector is excited by a suitably designed feed, the reflected wave is sensibly planar.[61]

However, unless the edges of the reflector are carefully designed, the incident field in the target zone will be contaminated by undesired fields diffracted from the edges of the reflector. The diffraction causes ripples in both the amplitude and the phase of the field distribution in the target zone. In some cases, the effect is small enough to be ignored, but in high-quality installations, the ripple may be objectionably large. Rolled-edge configurations, such as the one on the upper edge of the main reflector in Figure 14.24, can be designed to minimize edge diffraction.[62,63] However, the price paid for this improvement in performance is a much larger and more complicated reflector structure.

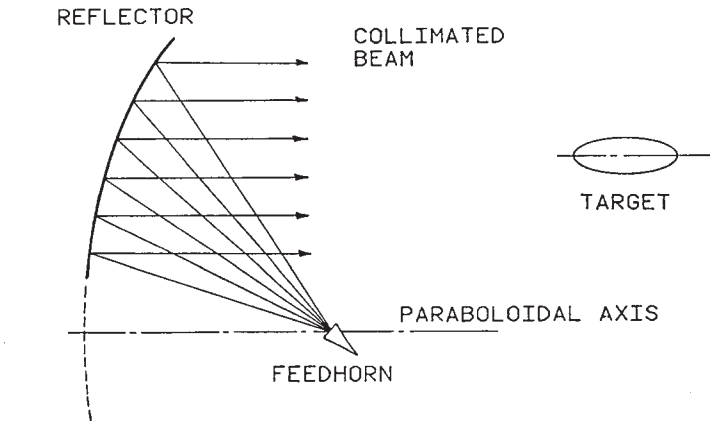


FIGURE 14.23 A compact range using an offset paraboloidal reflector

The single-reflector design shown in Figure 14.23 worked tolerably well for small test targets, but it soon became apparent that the reflector had to be considerably larger for many targets of vital military interest. This could be accomplished by doubling or tripling the size of the reflector, but that brought on a round of other problems, mainly the increased focal length of the reflector. While the focal distance could be shortened to comfortably fit axial space limitations (along the boresight of the reflector), it was then harder to control the field amplitude taper across the reflector due to the proximity of the feedhorn. The solution was to add a smaller subreflector in a subterranean gallery built especially for it, as suggested in Figure 14.24.[63,64]

The subreflector had the effect of greatly increasing the focal length of the main reflector, making it easier to optimize its illumination. This particular configuration is known as a *Gregorian system*, characterized by a focus between the two reflectors

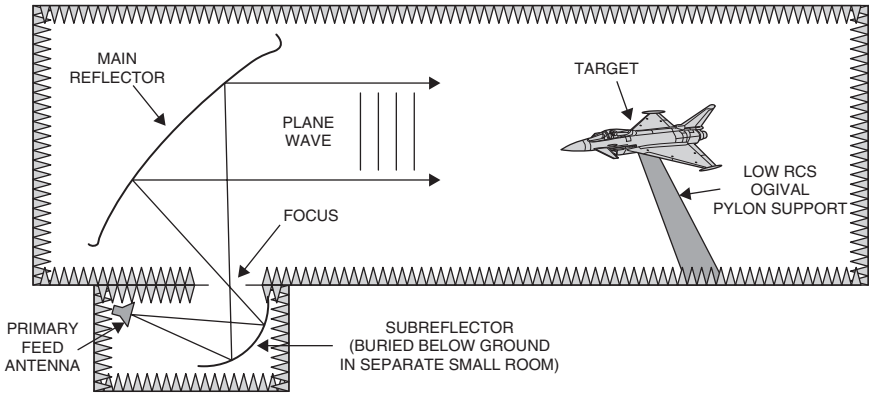


FIGURE 14.24 Generic dual-reflector compact range configuration[20]

where most of the rays between the two reflectors converge. In a second configuration known as the *Cassegrain*, the focus is a virtual one lying behind the subreflector. The Gregorian configuration permits the construction of a smaller aperture between the main chamber and the subreflector gallery than does the Cassegrain configuration.[65]

Large compact ranges such as the one shown in Figure 14.24 are very costly and can be afforded by only the largest of companies (or the government), thus range time is premium. Some companies keep their compact ranges busy 24 hours per day.

Targets measured indoors are placed much closer to the radar than those measured outdoors, and useful measurements may be made by using much less radiated power. Early indoor instrumentation radars relied on simple CW sources, and undesired chamber reflections were suppressed by a cancellation process. The procedure is to prepare the chamber for a measurement in every respect except for the installation of the target on its support fixture. A small sample of the transmitted signal is passed through a variable attenuator and a variable phase shifter and combined with the received signal. The amplitude and phase of the signal sample are then adjusted so as to cancel the signal received in the absence of the target.

The availability of low-cost, phase-locked, frequency-synthesized sources now makes it attractive to collect wideband RCS data, which contain far more target-scattering information than CW measurements made at single frequencies. When coherent RCS scattering data are suitably processed, it is possible to generate *ISAR imagery*, or two-dimensional maps of the echo sources of test objects.[66]

Figure 14.25 is an example of such an image. In this case, the processing required to generate this image is a double Fourier transformation, one from the frequency domain to the time domain and the other from the angle domain to the cross-range domain. The frequency-time domain processing may be performed virtually in real time (a second or two for processing and display on a video screen), but the conversion

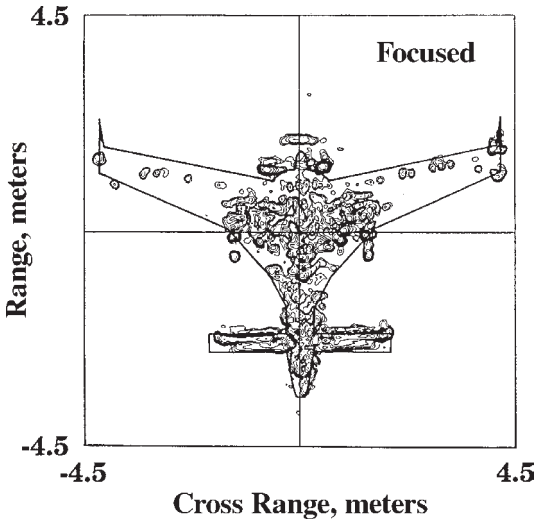


FIGURE 14.25 Radar image of a small drone aircraft. Planform outline has been added for clarity. (*Courtesy D. L. Mensa, U.S. Navy Pacific Missile Test Center.*)

from the angle domain to the cross-range domain must be performed offline in this example. The fast Fourier transform (FFT) is invariably exploited to expedite the processing.

The image data presented in Figure 14.25 were collected for a stepped-frequency signal with a bandwidth of 2.1 GHz centered at 3.4 GHz. The aspect angle was varied over a sector 35° wide centered on the nose-on aspect. The charted image contours are 5 dB apart and the total amplitude variation is 30 dB.

The resolution of the processed image in the time (range) domain is inversely proportional to the bandwidth of the emitted waveform. The resolution in the cross-range domain is inversely proportional to the aspect angle window over which the data are collected. Thus, the operating characteristics of the instrumentation system and the azimuthal data sampling rate must be decided before the data are collected. Because the cross-range coordinate of the resulting image is perpendicular to the axis of rotation of the target, it may be necessary to multiply that coordinate by a scale factor that effectively registers the generated image with, say, a plan view of the target.

The resulting data may be presented in the form of a contour map, as in Figure 14.25, or in a color-coded or grayscale pixel format. Here, the target outline has been superposed on the image data for diagnostic analysis, and the particular attitude shown is for nose-on incidence. Similar images can be generated for other angles of incidence, provided the target is rotated through a sector wide enough to yield the desired cross-range resolution and sampled at a sufficient number of angles over that sector. In practice, the target is rotated continuously while the swept- or stepped-frequency data are collected. A rule of thumb is that the angular speed be slow enough that the phase of the return at the end of a frequency sweep due to target motion not be more than 22.5° from what it would have been had the target not moved.

Note that the nose section of the drone is speckled with heavy concentrations of scattering centers, possibly due to contributions from internal structural features. The rear edges of the forward canards seem to be stronger scatterers than the leading edges. At the nose-on aspect angle for which the image data were collected, the leading edges of the wings are virtually invisible. However, note a few echo sources along a line parallel to, and slightly forward of, the wing trailing edges. If the target rotation had been centered on an aspect angle perpendicular to one of the wings, the leading edge of that one wing would have "lit up."

In the region of the main wing roots, we see heavy concentrations of echo sources. Some of them lie forward of any wing surface. Although we may conceive of apparent sources being strung out behind any physical scattering obstacle due to time delays of multiple reflections, it is hard to reconcile them being strung out *in front* of the body. We do see several clusters of apparent scattering centers positioned aft of the tail of the aircraft, but lacking any detailed description of the test object, we cannot interpret their meaning.

These "ghost" scatterers owe their existence to the way in which the data-processing system sorts the range and cross-range locations of scatterers. Down-range locations are sorted according to their processed time delays and cross-range locations according to their time-delay rates, whether due to real scatterers or to interactions between scatterers. Even though the contributions of some scattering centers may involve propagation in directions other than along the line of sight from the radar, the system has no way of discerning that fact. Therefore, despite the powerful diagnostic value of images like these, one must always be aware that multiple interactions between target elements can create scattering sources that are not where they appear to be.

14.5 RADAR ECHO SUPPRESSION

The probability of being detected by hostile radars can be reduced by decreasing the target's radar cross section. The major methods to reduce the RCS are by shaping of the target and the use of absorbers. By shaping, we mean the intentional selection of target surfaces and features so as to minimize the amount of energy scattered back to the radar. Shaping includes specific design configurations, such as the placement of engine intakes, which can have large radar echoes, where they may be shielded from the incident wave by other parts of the target. The purpose of radar-absorbing materials (RAM) is to absorb the incident radar energy so as to minimize the energy scattered back to the radar. Both methods have advantages and disadvantages.

Radar Absorbers. The purpose of the radar absorber is to attenuate incident energy and thereby reduce the energy scattered or reflected back to the radar. Most absorbers are designed to reduce specular reflections from metallic surfaces, but stealth technology has spawned the development of nonspecular absorbers intended primarily to suppress echoes due to surface traveling waves.

The simplest specular absorber is the *Salisbury screen*, which is a thin resistive sheet mounted a quarter wavelength above the metal surface to be hidden from a radar.[67] The design works best for incidence normal to the sheet, and if it can be manufactured with a resistivity of 377 ohms per square (the impedance of free space) all the power in the incident wave is transferred to the sheet and none is reflected. However, the single-sheet Salisbury screen suffers several limitations. Its thin sheets and low-loss spacers are fragile, its 20-dB bandwidth is barely 25%, and its performance deteriorates progressively as the angle of incidence moves away from normal incidence.

It is difficult to overcome the fragility problem with robust materials, but the bandwidth can be improved by cascading several sheets, as shown in Figure 14.26. This creates what is known as a *Jaumann absorber*. The bandwidth rises with the number of layers and can attain a respectable 140% for a four-sheet design, as suggested in Figure 14.27. The price paid for this bandwidth expansion is a much thicker, bulkier material that tends to be impractical for tactical military targets.

Like the Salisbury screen, the *Dällenbach layer* is also a simple absorber. The material is uniform throughout its volume and is a mixture of compounds designed to have a specified index of refraction. That design may include materials with magnetic losses as well as carbon particles responsible for electric losses. Therefore, the electric and magnetic susceptances (relative permittivity and relative permeability) have imaginary components, resulting in an index of refraction with an imaginary component.

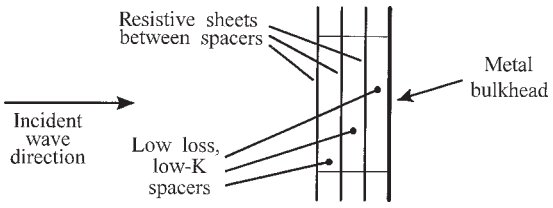


FIGURE 14.26 The Jaumann absorber is a cascaded collection of thin resistive sheets stacked in front of a metal bulkhead, spaced $\lambda/4$ apart, where K = spacer dielectric constant. The classic Salisbury screen is the degenerative case of a single sheet.

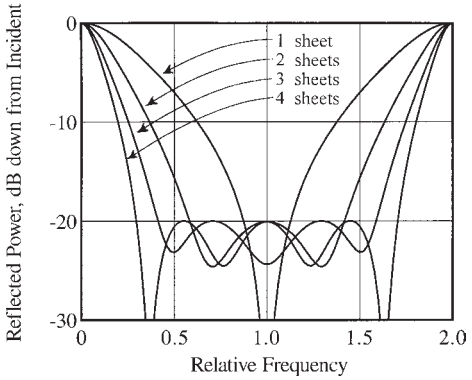


FIGURE 14.27 Performance of Jaumann absorbers with as many as four sheets. All four of these traces were optimized for maximum bandwidth at the –20 dB level. The sheet resistivities must increase from a low value at the inner sheet to a high value at the outer sheet.

The resulting imaginary part of the propagation constant attenuates waves traveling through the material.

Most of the commercial versions of Dällenbach absorbers are flexible and can be applied to modestly curved surfaces. The dielectric absorbers are typically made of a rubbery foam, sometimes urethane, impregnated with carbon particles. Impregnation may be performed by dipping a compressed slab of material in a graphite suspension bath, then wringing it out and drying it. Magnetic Dällenbach layers can be rolled from a mixture of natural or synthetic rubber loaded with carbonyl iron or ferrite powders. The lower the powder content, the more flexible the sheet but the less effective its electromagnetic performance. Dielectric and magnetic Dällenbach layers typically range from about 1 mm to 1 cm in thickness. The magnetic versions are as heavy 320 lb/ft^3, making them impractical for most tactical applications.

The front face and the metal backing of the Dällenbach layer are its only sources of reflection. The use of physically realizable materials makes it impossible to force either reflection to zero. The design objective, therefore, is to choose the electrical properties of the layer so that the front-face and metal-back reflections tend to cancel each other. If the material properties are dominated by electric effects, the optimum layer thickness is close to $\lambda/4$, as measured in the material. If they are dominantly magnetic, the layer can be much thinner.

As with the simple Salisbury screen, Dällenbach layers can be cascaded in attempts to expand bandwidth, producing what is known as *graded absorbers*. For optimum performance, the intrinsic impedance of each layer typically gets smaller the closer the layer is to the metal bulkhead or backing layer. Five or more layers have been used in the commercial production of graded dielectric absorbers, but commercial-graded magnetic absorbers appear to have been limited to three layers. It is important in the design process to account for the actual thickness and electrical properties of the adhesive films used to bond the layers to each other. These materials are too flimsy or too heavy for most military applications.

The pyramidal absorber used to suppress wall reflections in indoor chambers represents a particularly effective method of varying the effective impedance “seen”

by an incident wave. The absorber is made of flexible, carbon-impregnated plastic foam cut in the form of pyramids. It exhibits optimum performance when the pyramids are pointed toward the direction of the incident wave, and the pyramids should be of the order of 3 to 6 wavelengths deep. Fire-retardant paint is usually applied to pyramidal absorbers to satisfy safety requirements, but at high frequencies, the paint tends to degrade the performance of the material. Nevertheless, pyramidal absorbers of sufficient depth consistently turn in performances better than (less than) −50 dB. Because these absorbers do not rely on the cancellation of a front-face reflection by a rear-face reflection, they exhibit great bandwidth. In general, a pyramidal absorber with sharp tips and uniform bulk loss characteristics can have a bandwidth that exceeds 100:1.[60]

Nonspecular absorbers need not have the great thickness characterized by specular absorbers. Intended primarily for suppression of surface traveling-wave echoes, nonspecular materials have the opportunity to reduce the buildup of surface currents over several wavelengths *along the surface*. They are able to register quite respectable performances, therefore, simply because a thin layer attached to a metallic surface need not be very heavy. In this respect, the surface traveling-wave contribution due to long, smooth surfaces is one of the easiest to suppress. Even so, the thickness and geometric distribution of surface-wave absorbers should be varied for optimum performance.

By this point, it should be apparent that the application of radar absorbing materials to vulnerable targets is not a very effective way to enhance their survivability. These materials are heavy, demand undue surface care and maintenance, suffer limited bandwidth, and not least, are costly. A far more viable option, target shaping, is usually available if one is willing to consider it at the onset of system development.

Shaping. Shaping is the result of judiciously orienting target surfaces and edges in a way that minimizes their contributions to the total radar echo. This often means selecting airframe shapes and naval hull profiles that initially outrage airframe designers and naval architects. But there is no point in considering shape control unless a specific threat direction can be identified in azimuth or elevation or both. If all directions are equally likely, then the advantage of choosing favorable surface orientations for one threat direction is canceled by an accompanying enhancement in another. In many situations, however, the general direction of the threat can be forecast.

Figure 14.28 illustrates the RCS reduction available through shaping. The plotted curves are based on theory and measurements and show how the nose-on (axial) RCS varies with the electrical size of each of the six rotationally symmetrical metallic bodies shown in Figure 14.29. The diameters and projected areas of the objects are identical, and their volumes differ at most by a factor of 2. Except for the sphere, whose geometric-optics RCS is shown by the uppermost trace, all the objects have the same nose angle (40°), and of the six shapes, the ogive exhibits the lowest RCS. Thus, at least along the axes of these particular bodies, the RCS can be minimized by selecting the appropriate surface profile.

However, the attainment of low echoes over a range of aspect angles is usually accompanied by higher echo levels at other angles. Thus, the selection of an optimum shape should always include an evaluation of the variation of the RCS over a range of aspects wide enough to cover the anticipated threat directions. This implies the capability to measure the RCS patterns of a collection of objects with candidate surface profiles or the capability to predict those patterns, or both.

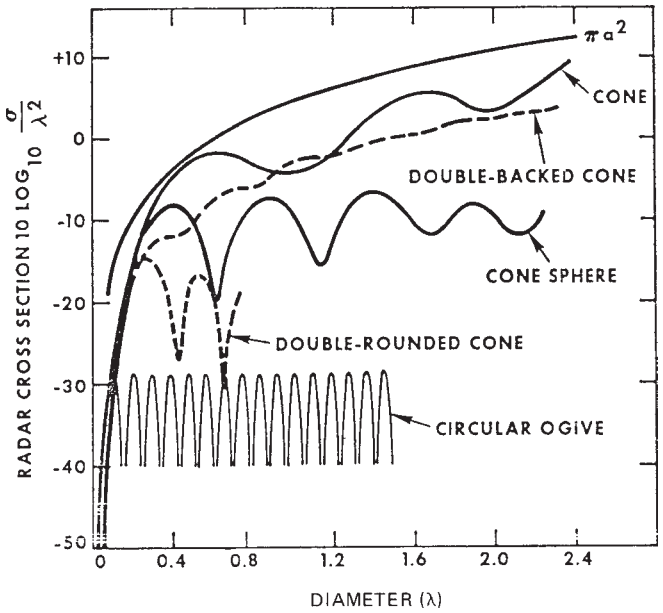


FIGURE 14.28 RCS of a collection of bodies of revolution of similar size and projected area (*after W. E. Blore*[68] *© IEEE 1964)*

Two approaches may be taken in the application of shaping. One is to replace flat surfaces with curved surfaces and thereby eliminate narrow but intense specular lobes. This is not very effective because it increases the general echo levels at nearby aspect angles. The second approach is to extend flat and singly curved surfaces so as to further narrow the specular lobe even if this increases its intensity. The logic of this approach is that the probability of detection is proportional to the average RCS over a range of solid angles of observation, and if the width of the lobe is narrow enough, its contribution to the average RCS may be less than if it were a wider but less intense lobe. The required RCS pattern levels of specific vehicle concepts should be established by means of mission analyses before deciding which shaping criterion is applicable.

Shaping is usually difficult to exploit or expensive to implement for vehicles or objects already in production because the vehicle configuration and profile have been selected and optimized for specific mission objectives. Changes in the configuration after production, therefore, are likely to impair the mission capabilities of the vehicle. If considered an option in the control of RCS, shaping must be included in the conceptual design of the vehicle well before any production decisions are made. Furthermore, shaping is not very effective for bodies that are not electrically large.

Low Radar Cross Section Vehicles. The following are some examples of low cross section vehicles from the early days of low cross-section vehicle deign.

The SR-71. The design of low cross-section vehicles probably began when Clarence (“Kelly”) Johnson, Lockheed’s famous designer, undertook development

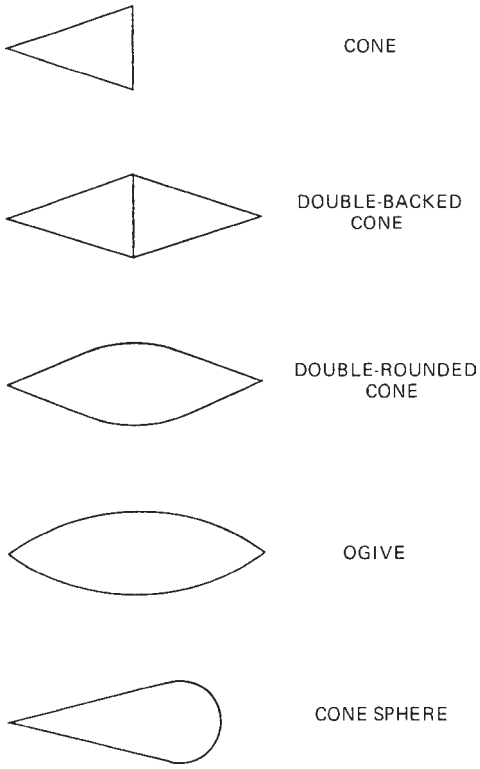


FIGURE 14.29 The objects whose radar cross sections are plotted in Figure 14.28.

of the A-12, an early prototype of the SR-71 Blackbird reconnaissance platform. At Lockheed's fabled Skunk Works, he recognized the importance of bringing the RCS design engineer into the inner-circle design team. The A-12 development contract was awarded in 1959,[69] and by early 1964 the A-12 was flying missions around the world.

The influence of Johnson's radar signature specialists is evident in Figure 14.30. The most remarkable feature of the SR-71 is the chine that runs from the nose to the root of the delta wing. In the nose-on sector, the radar echo will be dominated by returns from the engine inlets where the chine contribution is no doubt negligible. When seen from the broadside aspects, the chine reduces the specular echo that would have come from the otherwise rounded sides of the forward fuselage. Note that the tail fins are canted inward, thereby deflecting incident radar waves upward (away from the radar) when seen from the side. This design greatly reduces the SR-71's echo in a narrow broadside sector of angles.

The F-117. The largest echo source in the SR-71 nose-on region is probably the engine intakes because they are thrust well forward of the wing leading edges. The intakes of the F-117 Nighthawk, by contrast, are installed above the wing and well aft of the leading edge; they are the little black diamonds seen in the nose-on view of the F-117 shown in Figure 14.31. As such, the intakes are shielded from ground-based

FIGURE 14.30 Lockheed's SR-71 Blackbird

radars when the F-117 is in level flight. Even so, the F-117 developers contrived an egg-crate covering for the intakes that, in theory, prevents approaching radar waves from entering the inlet ducts, where they rattle around and eventually come back out again, heading directly back to the radar.[70]

Because the egg-crate grillwork reduced airflow, the intake had to be enlarged to restore normal flow. And the designers later discovered that the grillwork tended to ice up, which they overcame with an electrical heating system. The most remarkable features of the F-117 are its facetted (aerodynamically poor) surface profile and its

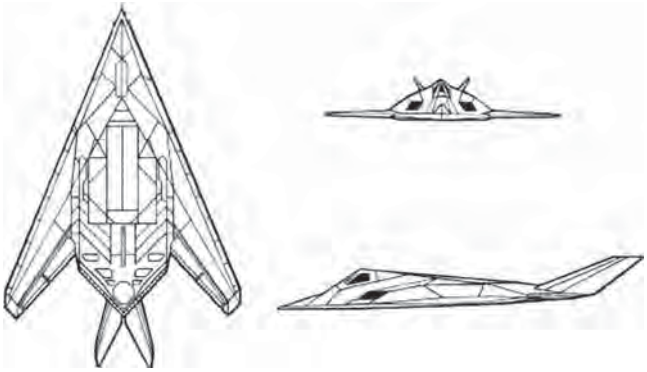


FIGURE 14.31 Lockheed's F-117 Nighthawk fighter

sharply swept tail fins. The tail fins are canted out as well as being swept aft, and the wings are also sharply swept. Although Lockheed managed to produce a supersonic SR-71 airframe using exotic materials, the low RCS demanded of the F-117 required a subsonic airframe. The airframe is covered with thin absorbing material whose edges are serrated to reduce reflections where hatches and covers fit into the fuselage.

B-2 Spirit Bomber. Northrop, the prime contractor for the B-2 Spirit, faced the same problem with engine placement that Lockheed did with the F-117 Nighthawk. The solution was the same: embed the engines' intakes in the top side of the airframe. Possibly because the B-2 had a different mission than the F-117 did, Northrop did not install egg-crate grillwork over the engine intakes. Indeed, Northrop designers devised hinged cowlings that open up at landing and takeoff speeds to increase airflow into the engines. The cowlings are then retracted at cruising speed for optimum thrust and efficiency.

In the interest of lowering the airframe's radar echo, Northrop's design team deliberately built the B-2 without a tailfin. Aside from this, it is interesting that the F-117 and B-2 share some common features. One is the rounded wingtips, which tend to reduce tip-spawned reflections of surface traveling waves that might build up along the wing's leading edges. Another is the use of thin absorber coatings that are serrated around the edges of hatches, covers, and canopies to suppress reflections from edge discontinuities. Yet another is the fact that the B-2, like the F-117, has a subsonic airframe. This suggests that although the chine concept may have worked for the fuselage of the fabled Mach-3 SR-71, it is by no means a viable approach for reducing the echoes from wing leading edges.

Indeed, the designers of the B-2 airframe apparently recognized early in the design work that it would be nearly impossible to reduce the leading edge RCS of their subsonic airframe to acceptable levels. That being the case, they decided to angle all leading and trailing edges at a common sweep angle: 34°. Their philosophy was that if they couldn't suppress the edge returns, the next best option would be to angle them all into four common directions in space. Thus, all trailing edges in the B-2 are parallel to leading edges.

X-45C Unmanned Combat Vehicle. The Boeing Company unveiled a mockup of the "C" version of its unmanned combat vehicle in the mid-2000s, less than a decade after the project was funded by the Defense Advanced Research Projects Administration (DARPA). An Associated Press photo of the aircraft mockup is shown in Figure 14.32. The airframe is 49 ft wide and 39 ft long.

The nose angle of the planform is nearly double that of the F-117, yet barely half that of the B-2, so the planform nose angle seems to be a compromise between the two. Note that the absence of a cockpit improves the stealthiness of the airframe. Indeed, the engine inlet cowling replaces the cockpit, and even the lips of the inlet are broadly serrated to reduce their contribution to the radar echo. The lack of any tailfin and the contour of the trailing edges of the planform reveal strong influences from the B-2 design philosophy.

Low Cross Section Ships. Conventional ships usually have very large structures with many horizontal and vertical reflecting surfaces that meet at right angles, forming corner reflectors with strong echoes. They also can have many individual scatterers that can contribute to a large radar cross section. To achieve low radar cross section, the horizontal/vertical design plan is changed so as to employ tilted surfaces

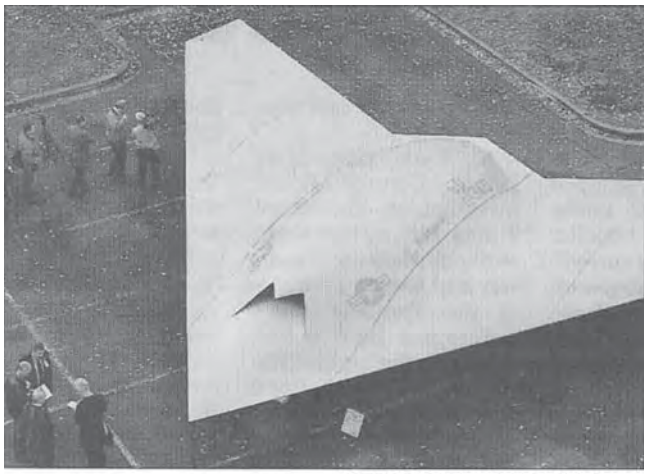


FIGURE 14.32 Boeing's mockup of its X-45C unmanned combat vehicle (*Courtesy of the Associated Press*)

that spoil the corner-reflector effect. Surface tilting also moves the radar angle of incidence well into the far sidelobes of most surfaces on the vessel. Thus vertical bulkheads often are tilted inboard.[71]

REFERENCES

1. E. F. Knott, "Radar observables," in *Tactical Missile Aerodynamics: General Topics*, Vol. 141, M. J. Hemsch, ed., Washington, DC: American Institute of Aeronautics and Astronautics, 1992, Chap. 4.
2. E. G Schneider, "Radar," *Proc. IRE*, vol. 34, pp. 528–578, August 1946.
3. S. D. Robertson, "Targets for microwave navigation," *Bell Syst. Tech. J*., vol. 26, pp. 852–869, 1947.
4. J. A. Stratton, *Electromagnetic Theory*, New York: McGraw-Hill Book Company, 1941, pp. 414–420, 563–567.
5. J. Rheinstein, "Backscatter from spheres: A short-pulse view," *IEEE Trans*., vol. AP-16, pp. 89–97, January 1968.
6. D. Atlas, L. J. Battan, W. G. Harper, B. M. Herman, M. Kerker, and E. Matijevic, "Back-scatter by dielectric spheres (refractive index ~ 1.6)," *IEEE Trans*., vol. AP-11, pp. 68–72, January 1963.
7. E. F. Knott, A. W. Reed, and P. S. P. Wei, "Broadside echoes from wires and strings," *Microwave Journal*, pp. 102 *et seq*, January 1999.
8. S. S. Chang and V. V. Liepa, "Measured backscattering cross section of thin wires, University of Michigan, Rad. Lab. Rept. 8077-4-T, May 1967.
9. L. Peters, Jr., "End-fire echo area of long, thin bodies," *IRE Trans*., vol. AP-6, pp. 133–139, January 1958.
10. P. S. P. Wei, A. W. Reed, C. N. Ericksen, and M. D. Bushbeck, "Study of RCS measurements from a large flat plate," in *Proc. 27th AMTA Conference, Antenna Measurement Techniques Association Symposium*, Newport, RI, October 31, 2005, pp. 3–8.
11. E. F. Knott, "A tool for predicting the radar cross section of an arbitrary corner reflector," in *IEEE Publ., IEEE Southeastcon' 81 Conference*, 81*CH*1650-1, Huntsville, AL, April 6–8, 1981, pp. 17–20.

12. E. F. Knott, "RCS reduction of dihedral corners," *IEEE Trans*., vol. AP-25, pp. 406–409, May 1977.
13. W. C. Anderson, "Consequences of non-orthogonality on the scattering properties of dihedral reflectors," *IEEE Trans*., vol. AP-35, pp. 1154–1159, October 1987.
14. E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, Raleigh, NC: SciTech Publishing, Inc., 2004, p. 254.
15. R. G. Hajovsky, A. P. Deam, and A. H. LaGrone, "Radar reflections from insects in the lower atmosphere," *IEEE Trans*., vol. AP-14, pp. 224–227, March 1966.
16. F. V. Schultz, R. C. Burgener, and S. King, "Measurements of the radar cross section of a man," *Proc. IRE*, vol. 46, pp. 476–481, February 1958.
17. C. R. Vaughn, "Birds and insects as radar targets: A review," *Proc. IEEE*, vol. 73, pp. 205–227, February 1985.
18. J. R. Riley, "Radar cross section of insects," *Proc. IEEE*, vol. 73, pp. 228–232, February 1985.
19. L. N. Ridenour (ed.), *Radar System Engineering*, MIT Radiation Laboratory Series, Vol. 1, New York: McGraw-Hill Book Company, 1947, p. 76.
20. U.S. Air Force web site, December 2005, *http://www.wrs.afrl.af.mil/other/mmf/compres.htm*.
21. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill Book Company, 1980, p. 45.
22. M. I. Skolnik, "An empirical formula for the radar cross section of ships at grazing incidence," *IEEE Trans*., vol. AES-10, p. 292, March 1974.
23. S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio*, 2nd Ed., New York: John Wiley & Sons, 1960, pp. 272–273.
24. J. A. Stratton, Ref. 4, pp. 464–467.
25. R. F. Harrington, *Field Computation by Moment Methods*, New York: Macmillan Company, 1968.
26. R. E. Kleinman, "The Rayleigh Region," *Proc. IEEE*, vol. 53, pp. 848–856, August 1965.
27. J. W. Crispin, Jr. and K. M. Siegel, eds., *Methods of Radar Cross Section Analysis*, New York: Academic Press, 1968, pp. 144–152.
28. E. F. Knott, "A progression of high-frequency RCS prediction techniques," *Proc. IEEE*, vol. 73, pp. 252–264, February 1985.
29. E. F. Knott et al., Ref. 14, p. 192.
30. E. F. Knott et al., Ref. 14, pp. 194–195.
31. T. B. A. Senior, "A survey of analytical techniques for cross-section estimation," *Proc. IEEE*, vol. 53, pp. 822–833, August 1965.
32. I. J. Gupta and W. D. Burnside, "Physical optics correction for backscattering from curved surfaces," *IEEE Trans*., vol. AP-35, pp. 553–561, May 1987.
33. J. B. Keller, "Diffraction by an aperture," *J. Appl. Phys*., vol. 28, pp. 426–444, April 1957.
34. J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Am*., vol. 52, pp. 116–130, 1962.
35. R. G. Kouyoumjian and P. H. Pathak, "A uniform theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, pp. 1448–1461, November 1974.
36. J. J. Bowman, P. L. E. Uslenghi, and T. B. A. Senior (eds.), *Electromagnetic and Acoustic Scattering by Simple Shapes*, Amsterdam: North-Holland, 1969, p. 258.
37. P. Ia. Ufimtsev, "Approximate computation of the diffraction of plane electromagnetic waves at certain metal boundaries, Part I: Diffraction patterns at a wedge and a ribbon," *Zh. Tekhn. Fiz. (U.S.S.R)*, vol. 27, no. 8, pp. 1708–1718, 1957.
38. P. Ia. Ufimtsev, "Approximate computation of the diffraction of plane electromagnetic waves at certain metal boundaries, Part II: The diffraction by a disk and by a finite cylinder," *Zh. Tekhn. Fiz. (U.S.S.R)*, vol. 28, no. 11, pp. 2604–2616, 1958.
39. P. Ia. Ufimtsev, "Method of edge waves in the physical theory of diffraction," U. S. Air Force Systems Command, Foreign Technology Division Doc. FTD-HC-23-259-71, 1971. (Translated from the Russian version published by Moscow: Soviet Radio Publication House, 1962.)

40. E. F. Knott et al., Ref. 14, pp. 209–214.
41. K. M. Mitzner, "Incremental length diffraction coefficients," Northrop Corporation, Aircraft Div. Tech. Rept. AFAL-TR-73-296, April 1974.
42. E. F. Knott, "The relationship between Mitzner's ILDC and Michaeli's equivalent currents," *IEEE Trans*., vol. AP-33, pp. 112–114, January 1985. [In the last term of Eq. (15) in this reference, the dot preceding the minus sign should be deleted and β should be replaced by sin β; in Eq. (20), the sign of the first term on the right side must be reversed.]
43. A. Michaeli, "Equivalent edge currents for arbitrary aspects of observation," *IEEE Trans*., vol. AP-32, pp. 252–258, March 1984. (See also correction, vol. AP-33, p. 227, February 1985.)
44. A. Michaeli, "Elimination of infinites in equivalent edge currents, Part I: Fringe current components," *IEEE Trans*., vol. AP-34, pp. 912–918, July 1986.
45. A. Michaeli, "Elimination of infinities in equivalent edge currents, Part II: Physical optics components," *IEEE Trans*., vol. AP-34, pp. 1034–1037, August 1986.
46. F. A. Sikta, W. D. Burnside, T. T. Chu, and L. Peters, Jr., "First-order equivalent current and corner diffraction scattering from flat plate structures," *IEEE Trans*., vol. AP-31, pp. 584–589, July 1983.
47. R. B. Mack, "Basic design principles of electromagnetic scattering measurement facilities," Rome Air Development Center Rept. RADC-TR-81-40, March 1981.
48. R. B. Dybdal, "Radar cross section measurements," *Proc. IEEE*, vol. 75, pp. 498–516, April 1987.
49. R. G. Kouyoumjian and L. Peters, Jr., "Range requirements in radar cross section measurements," *Proc. IEEE*, vol. 53, pp. 920–928, August 1965.
50. E. F. Knott et al, Ref. 14, p. 461.
51. M. A. Plonus, "Theoretical investigation of scattering from plastic foams," *IEEE Trans*., vol. AP-13, pp. 88–93, January 1965.
52. T. B. A. Senior, M. A. Plonus, and E. F. Knott, "Designing foamed-plastic materials," *Microwaves*, pp. 38–43, December 1964.
53. E. F. Knott and T. B. A. Senior, "Studies of scattering by cellular plastic materials," University of Michigan, Rad. Lab. Rept. 5849-1-F, April 1964.
54. C. C. Freeny, "Target support parameters associated with radar reflectivity measurements," *Proc. IEEE*, vol. 53, pp. 929–936, August 1965.
55. E. F. Knott et al., Ref. 14, p. 471.
56. W. H. Emerson, "Electromagnetic wave absorbers and anechoic chambers through the years," *IEEE Trans*., vol. AP-21, pp. 484–490, July 1973.
57. L. Solomon, "Radar cross section measurements: How accurate are they?" *Electronics*, vol. 35, pp. 48–52, July 20, 1962.
58. W. H. Emerson and H. B. Sefton, Jr., "An improved design for indoor ranges," *Proc. IEEE*, vol. 53, pp. 1079–1081, August 1965.
59. H. E. King, F. I. Shimabukuro, and J. L. Wong, "Characteristics of a tapered anechoic chamber," *IEEE Trans*., vol. AP-15, pp. 488–490, May 1967.
60. R. B. Dybdal and C. O. Yowell, "VHF to EHF performance of a 90-foot quasi-tapered anechoic chamber," *IEEE Trans*., vol. AP-21, pp. 579–581, July 1973.
61. R. C. Johnson, H. A. Ecker, and R. A. Moore, "Compact range techniques and measurements," *IEEE Trans*., vol. AP-17, pp. 568–576, September 1969.
62. W. D. Burnside, M. C. Gilreath, B. M. Kent, and G. L. Clerici, "Curved edge modification of compact range reflector," *IEEE Trans*., vol. AP-35, pp. 176–182, February 1987.
63. R. C. Rudduck, M. C. Liang, W. D. Burnside, and J. S. Yu. "Feasibility of compact ranges for near-zone measurements," *IEEE Trans*., vol. AP-35, pp. 280–286, March 1987.
64. W. D. Burnside, C. W. Pistorius, and M. C. Gilreath, "A dual chamber Gregorian subreflector for compact range applications," *Proc. of the Antenna Measurement Techniques Association*, September 28–October 2, 1987, pp. 90–94.

65. E. F. Knott, *Radar Cross Section Measurements*, New York: Van Nostrand Reinhold, 1993, pp. 33–38.
66. D. L. Mensa, *High Resolution Radar Imaging*, Norwood, MA: Artech House, 1981.
67. W. W. Salisbury, "Absorbent Body for Electromagnetic Waves," U.S. Patent 2,599,944, June 10, 1952.
68. W. E. Blore, "The radar cross section of ogives, double-backed cones, double rounded cones, and cone spheres," *IEEE Trans*., vol. AP-12, pp. 582–590, September 1964.
69. M. D. O'Leary and Eric Schulzinger, *SR-71: Inside Lockheed's Blackbird*, Oseola, WI: Motorbooks International Publishers & Wholesalers, 1991.
70. B. Sweetman and J. Goodall, *Lockheed F-117A*, Oseola, WI: Motorbooks International Publishers & Wholesalers, 1990.
71. E. F. Knott, "RCSR Guidelines Handbook," Final Technical Report on EES/GIT Project A-1560-001," Engineering Experiment Station, Georgia Institute of Technology, April 1976.

Chapter 15
Sea Clutter

Lewis B. Wetzel

Naval Research Laboratory (retired)

15.1 INTRODUCTION

For an operational radar, backscatter of the transmitted signal by elements of the sea surface often places severe limits on the detectability of returns from ships, aircraft, and missiles, navigation buoys, and other targets sharing the radar resolution cell with the sea surface. These interfering signals are commonly referred to as *sea clutter*, or *sea echo*. The search for a useful understanding of this important radar contaminant began with the collection and analysis of clutter data from operating radars, with the goal of establishing the relationship between clutter signals and the parameters of both the radar and the sea environment. Much of the earliest work took place during WWII and can be found in one of the comprehensive series of RADLAB volumes documenting the radar research that was done at that time.[1] But most of the clutter data from this period, and even into the 1960s, was collected in bits and pieces from isolated experiments, often with poor, incomplete, or misleading descriptions of the sea surface.

It would seem a simple matter to refine these results by using instrumentation radars operating over the wide variety of radar and environmental parameters encountered in practice. But while the parameters relating to the radar system and its configuration, such as frequency, cell size, polarization, grazing angle at the surface, etc., can be specified, selecting and quantifying the environmental parameters is quite another matter. First, it has not always been clear which environmental parameters are important. For example, wind speed certainly seems to affect clutter levels, but the correlation of clutter with, say, a ship's anemometer readings is often inconsistent. And although the state of agitation of the sea surface (*sea state*) appears to have a strong effect, it is a subjective measure and its relation to prevailing local winds is often uncertain. Moreover, it has been found that the temperatures of the air and the sea surface can affect the way in which measured wind speed is related to the generation of clutter-producing waves, yet the importance of these effects were unappreciated over most of the history of sea clutter measurements, so air and sea temperatures were seldom recorded. Finally, even if the importance of an environmental parameter has been recognized, it is often difficult (or too expensive) to measure it accurately in the field under real sea conditions.

While many aspects of sea clutter thus remained frustratingly ill-defined, the earlier work did disclose some general trends, such as the tendency of average clutter signal strengths at low to intermediate grazing angles to increase with the

grazing angle and the wind speed (or sea state), and generally to be greater for vertical polarization and in upwind/downwind directions. (For other reviews of sea clutter and its history, see Skolnik,[2] Nathanson,[3] and Long.[4]) Sea clutter is, however, a complex phenomenon, presenting various faces depending on the way the radar views the scene. For example, it is commonly noted that when viewed on an A scope (signal amplitude versus range), the appearance of sea clutter depends strongly on the size of the resolution cell, or *radar footprint.* For large cells, it appears *distributed* in range and may be characterized by a surface-averaged cross section with relatively modest fluctuations about a mean value. As the size of the resolution cell is reduced, clutter increasingly appears to contain sequences of isolated target-like, or *discrete*, returns that vary in time. At the higher resolutions, the discrete returns tend to stand well out of the background, occurring for both polarizations but most clearly evident with horizontal polarization at small grazing angles. These isolated returns are called *sea spikes* and are a common clutter component in this radar operating regime. Quite clearly, an understanding of sea clutter in all its aspects will be a considerable undertaking. Fortunately, a close relationship between radar and oceanography has grown up in the fields of remote sensing, leading to the accumulation of a large amount of information, both experimental and theoretical, about how scattering at radar frequencies relates to oceanographic variables. In many ways, this information serves as the basis of much of our current understanding of sea clutter.

In modeling sea clutter, there is a difference between a *theory,* which relates the physical scattering properties of the sea surface to the received signal, and a *characterization,* which provides a description of the sea clutter data in terms of a statistical model (e.g., Rayleigh, lognormal, Weibull, and K-distribution) that, although sometimes suggestive of physical processes in the underlying scattering, is of greater direct interest to the radar system designer in providing detection probabilities and false alarm rates.

Historically, attempts to provide a theoretical explanation of the observed behavior of clutter signals trace essentially from the work pursued during World War II and described in the well-known MIT Radiation Laboratory book mentioned above.[1] Unfortunately, the scattering models developed during this period, along with most of those published over the following decade, failed to account in any convincing way for the behavior of sea backscatter. In 1956, however, Crombie observed that at high-frequency (HF) wavelengths (tens of meters) scattering appeared to arise from a resonant interaction with sea waves of one-half of the incident wavelength, i.e., to be of the Bragg type.[5] Reinforced by the theoretical implications of various small waveheight approximations and wave-tank measurements under idealized conditions, the *Bragg model* was introduced into the microwave regime by many workers in the mid-1960s.[6–8] This produced a revolution in thinking about the origins of sea clutter because it involved the sea wave *spectrum*, thus forging a link between clutter physics and oceanography in what became the field of *radio oceanography*. However, fundamental conceptual problems in applying the Bragg hypothesis in microwave scattering, along with its inability to address significant aspects of measured sea clutter, have led through the years to continuing inquiry into the physical origins of sea scatter and how best to model it.[9–14] This being the case, speculation about physical models will be kept to a minimum in the sections on the empirical behavior of sea clutter. The problem of modeling sea scatter will be discussed separately in a later section.

15.2 THE SEA SURFACE

Close observation of the sea surface discloses a variety of features describable as wedges, cusps, waves, foam, turbulence, and spray, as well as breaking events of all sizes and masses of falling water. Any or all of these might contribute to the scattering of electromagnetic waves responsible for sea clutter. The basic oceanographic descriptor of the sea surface, however, is the *wave spectrum*, which says little about the details of these features, but contains a great deal of information about the sea surface in general and is central to the application of the Bragg scattering hypothesis. Accordingly, some tutorial material describing the spectral characterization of the sea surface is included in this section, along with a brief discussion of surface events such as wave breaking and other surface effects thought to contribute to the production of *sea spikes*.

There are basically two types of surface waves, *capillary* and *gravity*, depending on whether surface tension or gravity is the dominant restoring force. The transition between one and the other takes place at a wavelength of about 2 cm, so the smaller capillary waves supply the surface fine structure while gravity waves make up the larger and most visible surface structures. Sea waves have their origin primarily in the wind, but this does not mean that the "local" wind is a particularly good indicator of what the wave structure beneath it will be. In order to arouse the surface to its *fully developed* or *equilibrium* state, the wind must blow for a sufficient time (*duration*) over a sufficient distance (*fetch*). That part of the wave structure directly produced by these winds is called *sea*. But waves propagate, so even in the absence of local wind, there can be significant local wave motion due to waves arriving from far away, perhaps from a distant storm. Waves of this type are called *swell*, and since the surface over which the waves travel acts as a low-pass filter, *swell* components often take the form of long-crested low-frequency sinusoids.

The Wave Spectrum. The *ocean wave spectrum* describing the sea surface appears in several forms. If the time history of the surface elevation is monitored at a fixed point, the resulting time series may be processed to provide a *frequency spectrum* $S(f)$ of the surface elevation, where $S(f)df$ is a measure of the *energy* (i.e., square of the waveheight) of the waves in the frequency interval between f and $f + df$. Wave spectra have been measured in the open ocean primarily for gravity waves down to wavelengths of about 1 m. Open-ocean measurements of capillary wave spectra are especially difficult to perform.[15,16]

For a *gravity wave*, the frequency f and the wavenumber K are related by the dispersion relation

$$f = (1/2\pi)(gK)^{1/2} \tag{15.1}$$

where g is the acceleration of gravity and $K = 2\pi/\Lambda$, with Λ being the wavelength. Although each individual gravity wave obeys this relation, the waves at a point on the sea surface could come from any direction; so they are characterized by a two-dimensional propagation *vector* with orthogonal components K_x and K_y, where the K to be used in Eq. 15.1 is the magnitude $K = (K_x^2 + K_y^2)^{1/2}$. The wavenumber spectrum associated with $S(f)$ is a function of the two components of K and is commonly written as $W(K_x, K_y)$. This is called the *directional wave spectrum* and expresses the asymmetries associated with winds, currents, refraction, isolated swell components, etc.

For a given source of asymmetry such as the wind, various parts of the spectrum will display different directional behaviors. For example, in a fully developed sea, the larger waves will tend to move in the direction of the wind while the smaller waves will be more isotropic. Directional spectra are more difficult to measure and are obtained by a variety of experimental methods, such as an array of wave staffs to measure surface heights over a matrix of points, a multiaxis accelerometer buoy, stereo photography, and even by processing radar backscatter signals. However, a frequency spectrum measured at a point can contain no knowledge of wave direction, so a wavenumber spectrum $W(K)$ is often defined in terms of the frequency spectrum $S(f)$ by the relation

$$W(K) = S(f(K))(df/dK) \tag{15.2}$$

with the relation between f and K given by Eq. 15.1. To account for the wind direction, $W(K)$ is sometimes multiplied by an empirical function of K and direction ϕ relative to the (up)wind direction.

Oceanographers have not always been in complete agreement about the form of the frequency spectrum. Nonequilibrium wave conditions, inadequate sampling times, poor ground truth, etc., can contaminate the data set from which empirical spectra are derived. However, by careful selection of data from many sources, ensuring that only equilibrium (fully developed) sea conditions were represented and the wind was always measured at the same reference height (usually taken as 10 meters), Pierson and Moskowitz[17] established a simple empirical spectrum that has proven popular and useful. It has the form

$$S(f) = Af^{-5}\, e^{-B(f_m/f)^4} \tag{15.3}$$

where g is the acceleration of gravity, and $f_m = g/2\pi U$, corresponding to the frequency of a wave moving with a velocity equal to the wind speed U; A and B are empirical constants. This spectrum is illustrated in Figure 15.1 for several wind speeds. The effect of increasing wind speed is simply to move the low-frequency cutoff to lower frequencies along the high-frequency f^{-5} asymptote. (It should be noted that most of the oceanographers' spectra are based on measurements at relatively low frequencies and so cannot be taken seriously at frequencies above about 2 Hz. Nevertheless, these spectral forms are often used up to 20 Hz or greater in predicting radar clutter under the Bragg hypothesis.)

Converting this frequency spectrum into an isotropic wavenumber spectrum through Eq. 15.2 results in a spectrum of similar form, only with a K^{-4} asymptote. Phillips[17] derived this asymptotic behavior on dimensional grounds, and a widely used simplification, obtained by replacing the smooth peak in Figure 15.1 by a sharp cutoff, is generally referred to as the *Phillips Spectrum*:

$$\begin{aligned} W(K) &= 0.005K^{-4} \qquad K > g/U^2 \\ &= 0 \qquad\qquad\quad K < g/U^2 \end{aligned} \tag{15.4}$$

where the cutoff wavenumber corresponds to the frequency f_m of the peak in Eq. 15.3. Opposed to this highly simplified form are increasingly complex spectra based on more careful empirical studies[18] as well as more sophisticated theoretical considerations.[19]

In discussing the characterization of the sea surface by its spectrum, it must be kept in mind that the spectrum is a highly averaged description of how the *energy* of the surface is distributed among the wavenumbers, or frequencies, of the waves present on it.

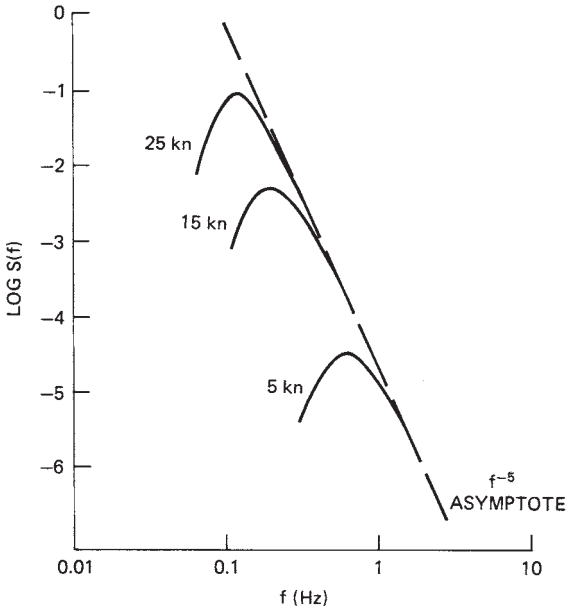


FIGURE 15.1 Sea wave frequency spectra of the Pierson-Moskowitz type, representing fully developed seas (*after W. J. Pierson and L. Moskowitz*[16] *© American Geophysical Union 1968*)

Since the phases of these waves are lost, the spectrum gives no information about the detailed morphology of the surface itself, i.e., about the complex surface features that are responsible for the scattered field. This point will be raised again as we go along.

General Sea Descriptors. The shape of the curves in Figure 15.1 indicates that the sea wave system is sharply peaked, so it should be possible to get a rough idea of the behavior of the major waves on the surface by taking the values of *period* $(1/f)$ and *wavelength* $(2\pi/K)$ defined at the spectral peak. These values are assigned to a wave satisfying the dispersion relation Eq. 15.1 and having a phase velocity $C = 2\pi f/K$ equal to the wind speed U. By using Eq. 15.1, the period T' and wavelength Λ' thereby defined take the form

$$T' = 0.64U \qquad \Lambda' = 0.64U^2 \tag{15.5}$$

where U is in meters per second. For example, the largest waves in a fully developed sea for a 15 kt (8 m/s) wind will have a wavelength of about 135 ft (41 m) with a period of 5 s.

The statistical distribution of waveheights on the ocean surface is quite close to gaussian, with a mean square height that can be obtained by integrating the wave-height spectrum over all frequencies (or wavenumbers). For spectra resembling those in Figure 15.1, the rms waveheight is given approximately by

$$h_{\text{rms}} = 0.005U^2 \text{ m} \tag{15.6}$$

The rms waveheight contains contributions from all the waves on the surface, but very often it is the peak-to-trough height for the higher waves that is of major interest. This is certainly the case for a ship in a seaway or in the shadowing of the surface at low radar grazing angles. The *significant height*, or peak-to-trough height of the one-third highest waves, provides such a measure. It is denoted by $H_{1/3}$ and is taken to be about six times the spectral rms amplitude (see, e.g., Kinsman,[20] Fig. 8.4-2). For a 15-kt wind, this is only about 2 ft, but for gale-force winds of 40 kt, it rises to almost 15 ft, which is a rather formidable sea.

Looking at the sea, an observer might describe what is seen in terms of a subjective *state of the sea*, e.g., "smooth," "rough," or "terrifying!" If these descriptions are listed in order of severity and assigned numbers, these numbers define a *sea state*. A similar numerical scale exists for wind speeds, the *Beaufort wind scale*, with numbers about an integer higher than the corresponding sea state. But it is seldom used in reference to sea clutter.

There are, then, two numbers commonly used to indicate the activity of the sea surface: a subjective sea state and a measured wind speed. Only when the wind has sufficient *fetch* and *duration* to excite a *fully developed* sea, can a waveheight be unambiguously associated with it. The surface descriptors generally used in connection with sea clutter—sea state, wind speed, and its associated equilibrium waveheight—are given in Table 15.1, with the wind speed in knots, the significant waveheight in feet, and the duration/fetch required for a fully developed sea in hours/nautical mile. It is of interest to note that the median wind speed over the world's oceans is about 15 kt, corresponding to sea state 3.

Breaking Waves and Other Surface Disturbances. The observable features of the sea surface that best suggest an origin for the sharp localized radar returns called *sea spikes* are surface events that are themselves sharply localized, events including breaking waves of all sizes, induced either by the wind or by nonlinear interactions among wave systems. Large-scale breaking waves display two characteristic behaviors—*spilling*, in which an unstable wave peak unravels, and *plunging*, where the peak curls over on itself and crashes onto the front face as a cascade of water masses, ending in a chaotic jumble.[20] Another different event is the *microbreaker*, a small, transient shock front induced by a puff of wind or another wave. As noted earlier, highly averaged wave spectra cannot disclose the morphology of such surface features, and, unfortunately, physical oceanography is still unable to provide a generally satisfactory description or characterization of wave breaking.[21] Nevertheless, there are two useful heuristic parameters relating elements of a breaking wave scene to wind speed. Whitecap density is a visible tracer of breaking wave activity and has a power-law dependence on wind speed

TABLE 15.1 Sea-Surface Descriptors

Sea State	Wind Speed, kt	Waveheight $H_{1/3}$, ft	Duration/fetch, h/nmi
1 (smooth)	< 7	1	1/20
2 (slight)	7–12	1–3	5/50
3 (moderate)	12–16	3–5	15/100
4 (rough)	16–19	5–8	23/150
5 (very rough)	19–23	8–12	25/200
6 (high)	23–30	12–20	27/300
7 (very high)	30–45	20–40	30/500

given by $\rho_{wb} \sim U^{3.5}$.[22] The average length of a breaking wavefront moving at speed c also depends on wind speed and is given by a parameter $\Lambda(c)$.[23] These parameters will appear again later when we discuss some of the more recent models for sea clutter. An additional feature of small-scale breaking, or other strongly nonlinear events, is the appearance of "parasitic" or "bound" capillaries attached to the event and moving with it.[20,24] They tend to be small-amplitude features, localized and narrow-band.

15.3 EMPIRICAL BEHAVIOR OF SEA CLUTTER

Sea clutter is a function of many parameters, some of them showing a complicated interdependence, so we emphasize again that it is not an easy task to establish its detailed behavior with a great deal of confidence or precision. For example, in a proper sea clutter measurement, the polarization, radar frequency, grazing angle, and resolution cell size will have been specified. Then the wind speed and direction must be measured at a reference altitude, and if the results are to be compared with those of other experimenters, the proper *duration* and *fetch* should be available to ensure standardization to equilibrium sea conditions. Since these measured winds are related to the wind structure at the surface through the atmospheric boundary layer, the shape of this layer must be determined by measuring the air and sea temperatures. To complicate the picture still further, it has been found that sea clutter can be dependent on the direction of the long waves, which includes *swell* in the measurement area, so ideally the *directional wave spectrum* should be measured as well. Obviously, it is unlikely that all of these environmental parameters will be recorded with precision in every (or even *any*) sea clutter measurement; so considerable variability in the basic conditions under which sea clutter data are collected by different experimenters can be expected. It is of interest to note that in many of the reported measurements of sea clutter, particularly in the older literature, wide inconsistencies between wind speed and waveheight may be found. For example, a wind speed of 5 kt might be reported with waveheights of 6 ft, or 20-kt winds with 2-ft waves. These pairings are not consistent with the values for an equilibrium sea described in Table 15.1 and indicate the unnoticed or unrecorded presence of heavy swell or highly nonequilibrium wind conditions or both. Even with all the variables properly specified, recorded clutter data can be spread over a wide dynamic range, especially at low grazing angles.

Since sea clutter is generally viewed as a surface-distributed process, the basic clutter parameter is taken to be the normalized radar cross section (NRCS), σ^0, of the surface, commonly referred to as *sigma zero* and expressed in decibels relative to 1 m^2/m^2. It is obtained experimentally by dividing the measured radar cross section of an illuminated patch of the surface by a normalizing area, so differences in the definition of this area can lead to inconsistencies among various reports of NRCS measurements. Scattering from any distributed target involves the product of the transmitting and receiving system footprints integrated over the target. These footprints cover exactly the same area for a monostatic radar and will depend on the pulse- and beamwidths, the range, and the grazing angle. If the footprints are assumed to be of the *cookie-cutter* type (constant amplitude falling sharply to zero at the half-power points), then the relation between the actual radar clutter cross section σ_c, as inferred from the received power via the *radar equation*, and the NRCS σ^0 is given by

$$\sigma^0 = \sigma_c / A_f \tag{15.7}$$

where for a radar with an antenna beamwidth B and rectangular pulse of length τ, viewing the surface at range R and grazing angle ψ, the area A_f is either

$$A_f = \pi (BR)^2/4 \sin \psi \tag{15.8}$$

for beam-limited conditions (e.g., continuous-wave (CW) or long-pulse radar at high grazing angles) or

$$A_f = (c\tau/2)BR/\cos \psi \tag{15.9}$$

for pulse-width-limited conditions (e.g., short-pulse radar at low grazing angles).

Real radars do not produce cookie-cutter footprints, however, since the antenna beam will have a complex profile and the pulse might be shaped. For this reason, an effective A must be obtained from a surface integration of the actual amplitude profile of the footprint, which will tend to result in a smaller value of A than that defined by Eq. 15.8 or Eq. 15.9. This will produce larger values of σ^0 as derived from measured values of σ_c by Eq. 15.7. Most experimenters use the half-power beamwidth in Eq. 15.8) or Eq. 15.9, with an error that is usually only 1 or 2 dB.

Clutter Statistics. Summaries of clutter measurements made before about 1970 may be found in several of the standard reference books on radar[2,3] and radar clutter.[4] Among the programs of this period, the most ambitious was that pursued in the late 1960s at the Naval Research Laboratory (NRL),[25] in which an airborne four-frequency radar (4FR), operating with both horizontal and vertical polarizations at UHF (428 MHz), L band (1228 MHz), C band (4455 MHz), and X band (8910 MHz), made clutter measurements upwind, downwind, and crosswind in winds from 5 to 50 kt for grazing angles between 5° and 90°. The system was calibrated against standard metal spheres dropped from aircraft, and wind speeds and waveheights were recorded in the measurement areas from ship instruments.

Typically, samples of σ^0 for a given set of radar and environmental parameters are scattered over a wide range of values and in the NRL measurements were organized into probability distributions of the type shown in Figure 15.2. The data, represented by the solid line, are plotted on normal probability paper with Rayleigh and log-normal distributions shown for comparison (dashed lines). The ordinate is the *percent of time for which the abscissa is exceeded*, and the abscissa is the value of σ^0 as defined by Eq. 15.7, with A taken from Eq. 15.8 or Eq. 15.9 as appropriate. This particular distribution is representative of clutter from a relatively large radar footprint (pulse length about 0.5 μsec or 75 meters) measured at intermediate grazing angles (20° to 70°) for moderate wind speeds (about 15 kt). It is Rayleigh-like but shows a tendency toward log-normal behavior for the larger cross sections. From a detailed statistical analysis of the NRL 4FR data, Valenzuela and Laing[26] concluded that, for this data at least, the distributions of sea clutter cross sections were intermediate between the Rayleigh and log-normal distributions.

Organizing the data samples into probability distributions makes the *median* (50%) value a convenient statistical measure of the clutter cross section. But many investigators process their data to provide the *mean* value, and because the conversion of a *median* to a *mean* requires knowledge of the probability distribution function, care must be taken to avoid ambiguity in comparing the measurements of different experimenters. The original analysis of the NRL 4FR data was based on *median* cross sections and the assumptions of the cookie-cutter antenna beam embodied in Eqs. 15.8 and 15.9.[25–27] In later presentations of this data,[28] the *median* values of σ^0

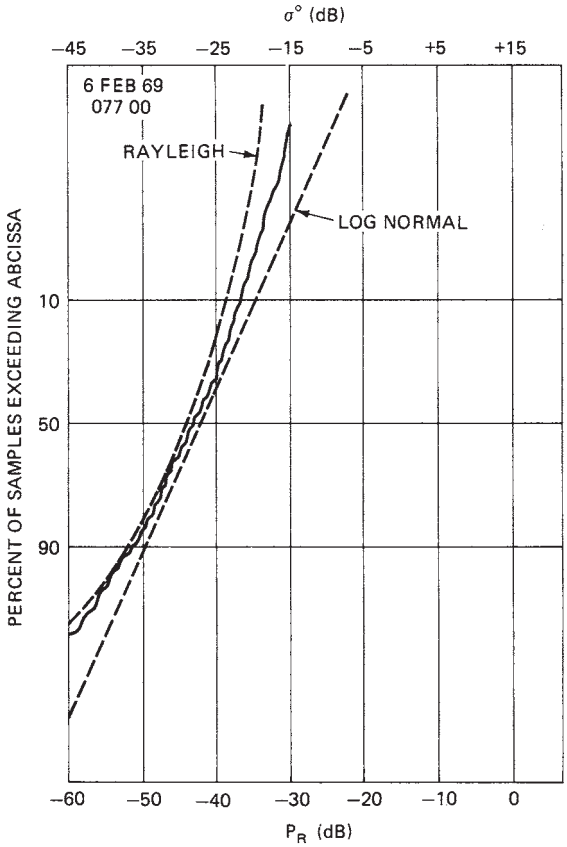


FIGURE 15.2 An example of the probability distribution of sea clutter data (*from J. C. Daley et al.*[25])

were replaced by *means*, raising them by about 1.6 dB, and the area A in Eq. 15.7 was redefined in terms of a more realistic tapered footprint, adding another 1 to 2 dB. This means that there can be a difference of 3 to 4 dB between the earlier and later presentations of the same data, and since these results have been widely used and quoted, it is important to ensure that the proper definition of σ^0 is being used when comparing them with clutter data that has been taken by other experimenters or in using these results in clutter predictions.

Figure 15.2 shows that even for intermediate grazing angles in the range 20°–70° the sea clutter distribution departs from strictly Rayleigh. At lower grazing angles, and particularly for narrow pulse widths, the presence of *sea spikes* or other non-gaussian behavior may be accommodated by one of the multiparameter or compound distributions that express an excess of higher returns, such as the Weibull and K-distributions. The latter was introduced to characterize the particular behavior of low-grazing-angle clutter seen in a marine environment.[29] Its success is very likely due to its relation to the Rice distribution, which describes the statistics of steady signals in noise, thus reflecting the statistics of "target-like" sea spike returns in a Rayleigh background.[30]

General Trends. Being the first really comprehensive collection of clutter data over a wide range of radar frequencies, the 4FR program produced many plots showing the dependence of sea clutter on grazing angle, frequency, polarization, wind direction, and wind speed. However, comparison of these plots with others made both earlier and later shows the extent of the variations to be found in sea clutter measurements reported by different investigators for exactly the same set of parameters. This is seen clearly in Figure 15.3*a* and *b*, which compares the grazing-angle dependence of X-band clutter data for wind speeds in the neighborhood of 15 kt obtained from four sources: NRL 4FR[28] (these are *mean* results for upwind directions and include the antenna corrections mentioned above), aircraft measurements by Masuko et al.[31] (also in the upwind direction), and summaries of data taken from books on radar systems by Skolnik[2] and Nathanson.[3] The discrepancies between the different data sets can be accounted for, at least in part, as follows. The older data summaries were based on published measurements from various sources in which there is no specification of wind direction. It may, therefore, be assumed that these data represent some kind of average of upwind, downwind, and crosswind directions. As will be seen, this average is about 2 to 3 dB smaller than the upwind returns. Moreover, the *early* NRL 4FR data was used liberally in the older data summaries, and it was noted above that there is a difference of 3 to 4 dB between the early and later presentations of the same NRL 4FR data, the *latter* being used in Figure 15.3*a* and *b*. With these corrections, the curves might show closer agreement. Nevertheless, it is clear that uncritical use of published clutter data could lead radar systems designers to choose sea clutter estimates many dB apart for the same conditions.

The NRL 4FR data set is unique in that no other program has reported measurements made over so wide a range of frequencies, grazing angles, and wind speeds at the same time. Figure 15.4 shows the *trends* for both vertically and horizontally polarized

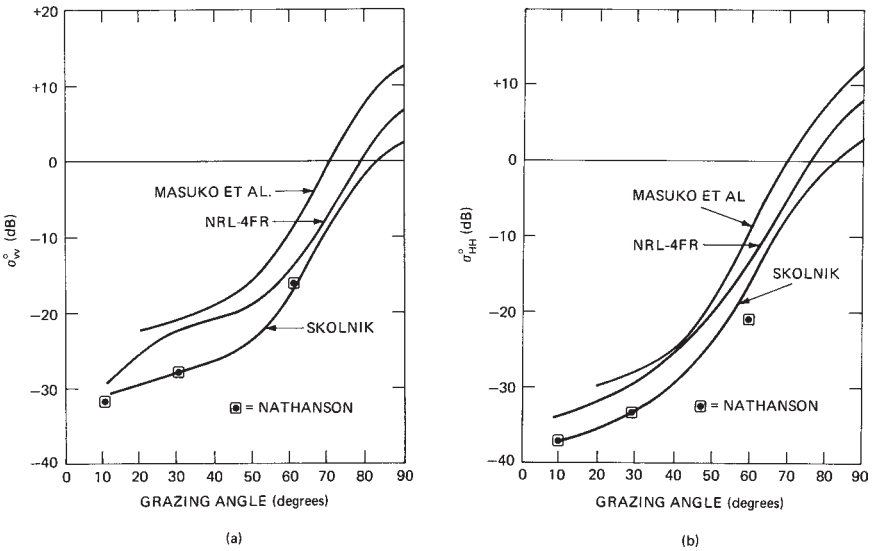


FIGURE 15.3 Comparison of X-band clutter data from different sources for a nominal wind speed of 15 kn: (*a*) vertical polarization and (*b*) horizontal polarization (*based on data from H. Masuko et al.,*[31] *NRL 4FR,*[25] *M. I. Skolnik,*[2] *and F. E. Nathanson*[3])

sea clutter over a range of grazing angles down to 5°. The curves represent the *centers* of ±5 dB bands that contain the major returns for the three higher frequencies (L, C, and X bands—the UHF returns were a few decibels lower) and wind speeds above about 12 kt. The major differences in sea clutter for the two polarizations are seen to lie in the range of grazing angles between about 5° and 60°, where the horizontally polarized returns are smaller. This difference is found to be emphasized at both lower wind speeds and lower frequencies. The cross sections approach each other at high angles (>50°) and, for the higher microwave frequencies, at low angles (<5°) as well. In fact, for grazing angles less than a few degrees and moderate to strong wind speeds, observers have reported that at X band and at the higher sea states the horizontally polarized returns can exceed the vertically polarized returns.[1,32,33]

The NRL 4FR system permitted transmission and reception on orthogonal polarizations so that data could be collected for cross-polarized sea clutter. These returns tended to have a weak dependence on grazing angle and were always smaller than either of the like-polarized returns, lying in the cross-hatched region shown on Figure 15.4.

It is informative to compare measurements at different frequencies by different investigators in different parts of the world under similar wind conditions. Figure 15.5 displays measurements of vertically polarized sea clutter down to a grazing angle of 20° for wind speeds of about 15 kt from three independent experiments using airborne radars at C-, X-, and K-band frequencies.[31,34,35] Although there is no assurance that

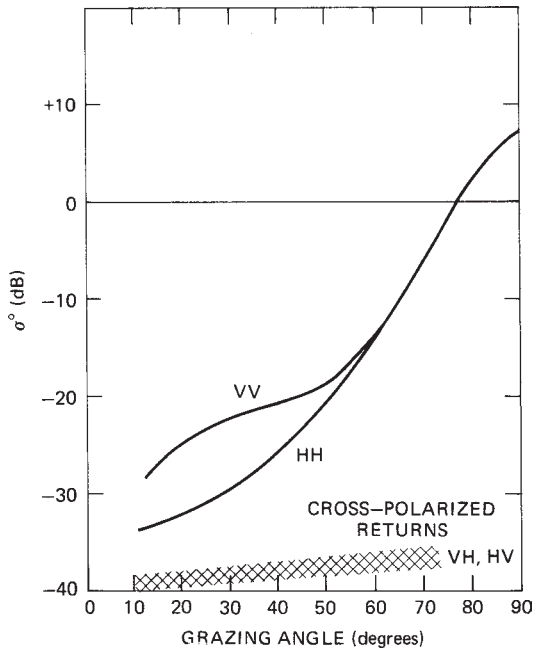


FIGURE 15.4 General trends in clutter behavior for average wind speeds (about 15 kt) based on NRL 4FR data. Plots represent L-, C-, and X-band data within ±5 dB.

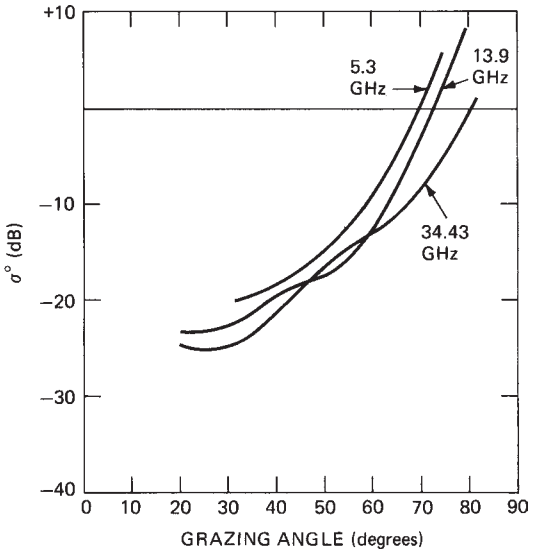


FIGURE 15.5 Frequency dependence of sea clutter for wind speeds of about 15 kt: 5.3 GHz, Feindt;[34] 13.9 GHz, Schroeder;[35] and 34.4 GHz, Masuko[31]

all these measurements were made over fully developed seas, it is clear that there is a rather strong consistency among them, which reinforces the observation made in reference to Figure 15.4 that the frequency dependence of sea clutter at intermediate grazing angles is weak at microwave frequencies from L to K band.

Dependence on Wind Speed and Direction. Experimentally, the relation between sea clutter and wind speed is complex and uncertain, it having been found to depend on almost all of the parameters that characterize sea clutter: frequency, grazing angle, polarization, the state of the sea surface, the direction and speed of the wind itself, and even on whether the measurements are made from an aircraft or a tower platform.[36]

A common way to organize clutter data is to seek the best straight-line fit (linear regression) between clutter cross sections in decibels and the log of the wind speed (or some other parameter). This, of course, *imposes* a power-law relation between the variables: $\sigma^0 \sim U^n$, where n is determined by the slope of the line. An example is shown in Figure 15.6.[37] On the other hand, while the totality of the NRL 4FR results appeared to show saturation for wind speeds above about 20 kt, the high and low-to-moderate wind-speed data were collected at different times in different places under different conditions of sea-surface development, and discrepancies between the two data sets for common wind speeds have weakened the evidence for saturation.[38] Other investigators deny that it is even possible to express wind dependence in the form of a power law, proposing the existence of a kind of threshold wind speed, below which clutter virtually vanishes and above which the clutter level rises rapidly toward a saturation value.[18] This is indicated by the curves in Figure 15.7, where the straight lines correspond to various power laws and the curved lines derive from

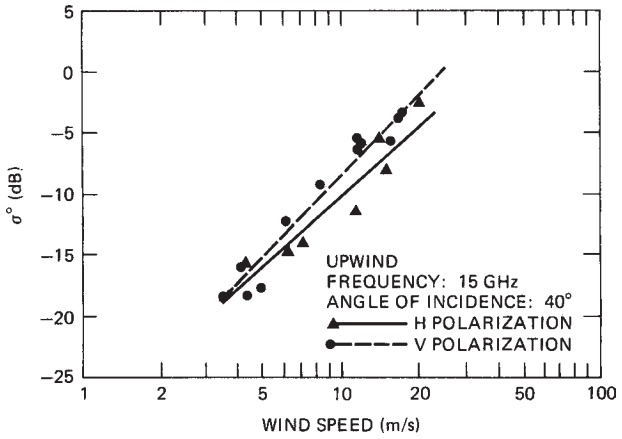


FIGURE 15.6 Sea clutter from a tower platform with power-law wind-speed dependence defined by linear regression (angle of incidence = 90° grazing angle) (*after A. H. Chaudhry and R. K. Moore,*[37] *© IEEE 1984*)

wave spectrum considerations.[18] It is possible to find examples of data that appear to follow such behavior while at the same time being expressed as a power law by brute linear regression, as illustrated in the tower data shown in Figure 15.8.[37] This behavior is not uncommon.

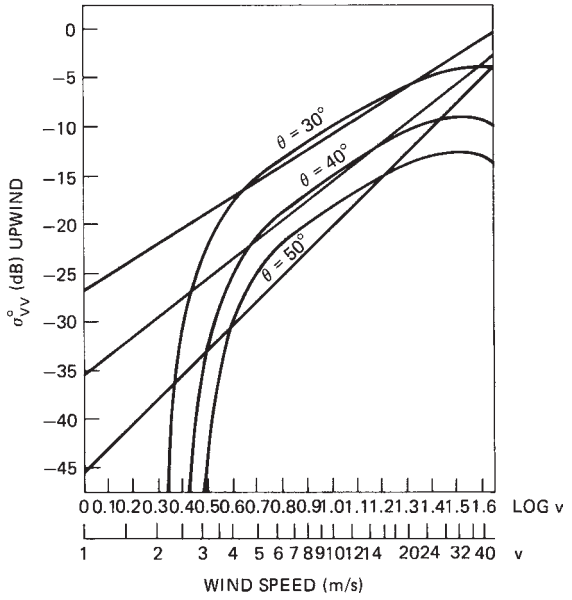


FIGURE 15.7 A hypothetical wind-speed dependence of sea clutter (curved traces) compared with various power laws (straight lines) (*derived from W. J. Pierson and M. A. Donelan*[18] *© American Geophysical Union 1987*)

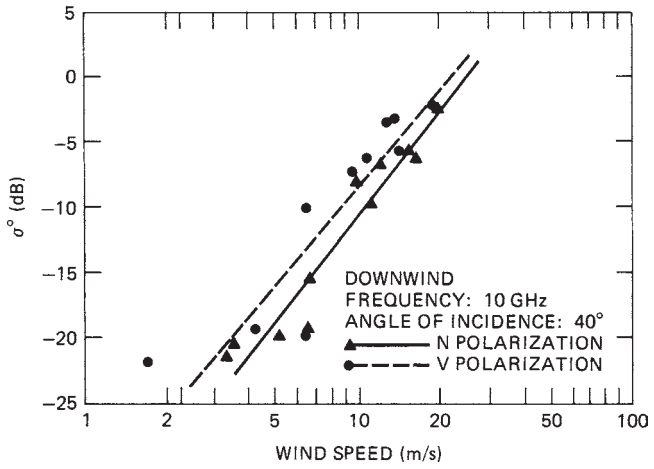


FIGURE 15.8 Example of forcing a power-law fit (compare data points with those in Figure 15.6) (*after A. H. Chaudhry and R. K. Moore*[37] *© IEEE 1984*)

Nevertheless, the imposition of a power-law relation provides a convenient way to *visualize trends* in the behavior of sea clutter with wind speed. The various aircraft measurements referred to above,[31,34,35] augmented by data from a tower in the North Sea,[36,37] were used as the basis of the power-law plots of σ^0 as functions of wind speed and grazing angle shown in Figure 15.9*a*,*b*. These plots suggest how sea clutter for a given frequency (X band), wind direction (upwind), and polarization behaves with wind speed and grazing angle. However, examination of the actual data points underlying these linear regressions show point scatter that sometimes resembles Figure 15.6, sometimes Figure 15.8, and sometimes neither, so these straight lines cover up considerable uncertainty. In fact, it appears that there is no simple functional dependence of sea clutter on wind speed that can be established with any confidence from existing data, although most investigators would probably agree that the behavior of microwave sea clutter with wind speed at intermediate grazing angles can be roughly described as follows: for light winds (less than 6 to 8 kt) sea clutter is weak, variable, and ill defined; for intermediate winds (about 12 to 25 kt), it can be described roughly by a power law of the type found in Figure 15.6; and for strong winds (above about 30 kt), there is a tendency for it to level off. In fact, the convergence of the lines in Figure 15.9*a*,*b* with increasing wind speed suggests that the reflectivity of the sea surface is tending toward Lambert's law, for which there is no dependence on grazing angle, frequency, or polarization but only on surface *albedo*, or average reflectivity.

In several of the experiments referenced above, the dependence of sea backscatter on angle relative to the wind direction was found by recording the radar return from a spot on the surface while flying around it in a circle. Figure 15.10*a*,*b* gives an example of this behavior for grazing angles of about 45° and wind speeds close to 15 kt.[31] The figures contain results obtained independently by three different groups. The behavior shown here is representative of that found generally: sea clutter is strongest viewed upwind, weakest viewed crosswind, and of intermediate strength viewed downwind, the total variation being about 5 dB. Other studies corroborate this behavior.[39]

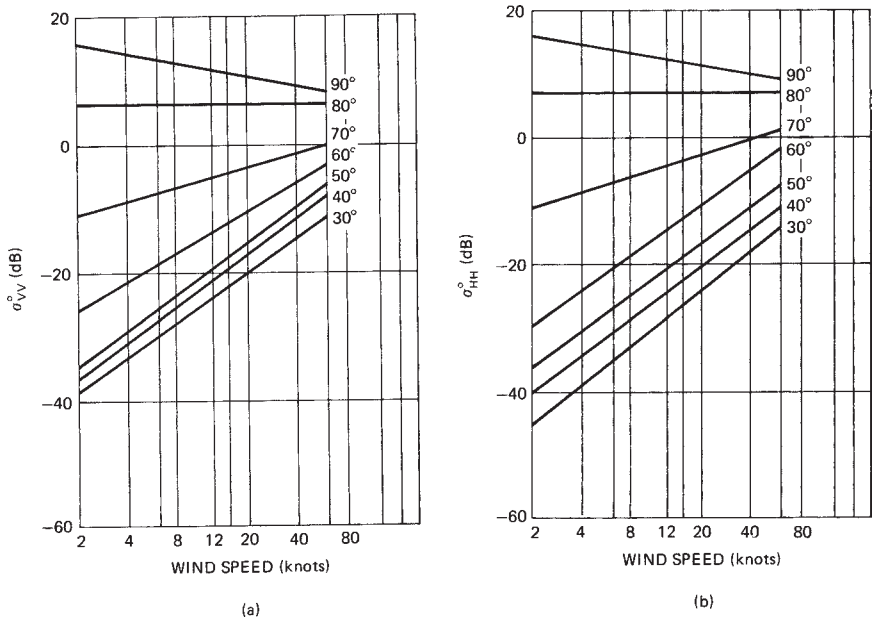


FIGURE 15.9 A representation of X band, upwind clutter behavior with wind speed and grazing angle: (*a*) vertical polarization and (*b*) horizontal polarization

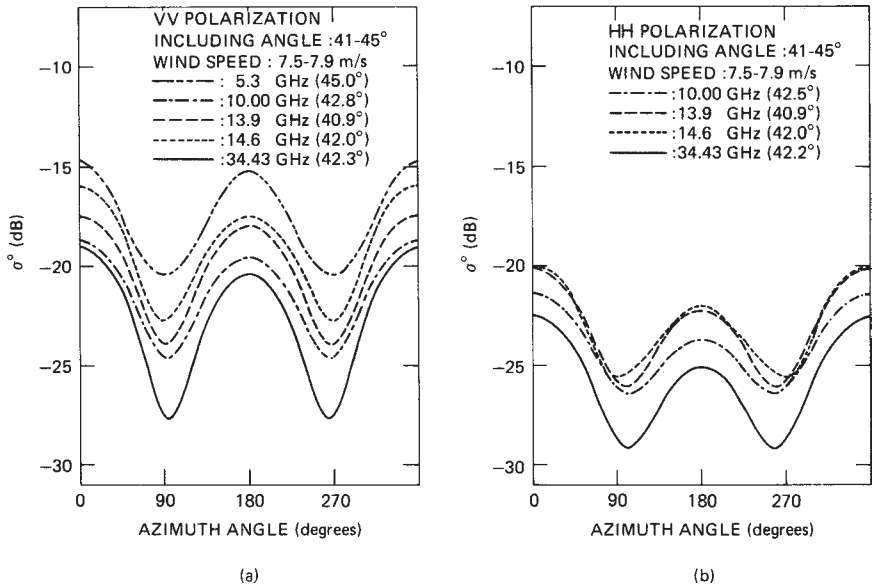


FIGURE 15.10 Dependence of clutter on wind direction: nominal wind speed, 15 kt; grazing angle about 45°; upwind, 0°, 360°; and downwind, 180° (*after H. Masuko et al.,*[31] *© American Geophysical Union 1986*)

Sea Clutter at High Grazing Angles. The top lines in Figure 15.9*a*,*b* correspond to clutter at a grazing angle of 90°, that is, for a radar looking straight down. On a strictly empirical basis, the clutter cross section at this angle is only weakly dependent on frequency, has a maximum of about 15 dB at zero wind speed (at least for the antenna beamwidths and experimental configurations reported), and falls off gradually as the wind picks up. Scattering at high grazing angles is commonly regarded as a form of specular scattering from tilted facets of the surface, so it is of interest to note that there appears to be a small range of angles in the neighborhood of 80° for which the cross section is almost completely independent of wind speed. Since these angles correspond to complements of the common rms sea slope angles of about 10°, it might be argued that as the wind increases, the clutter *decrease* due to increasing surface roughness is balanced by a clutter *increase* due to an increasing population of scattering facets. This line could, therefore, be regarded as the boundary separating the *specular* regime, where the cross section is decreased by surface roughness, from the *rough-surface* regime, where the cross section increases with surface roughness. It should further be noted that clutter measurements at these high grazing angles will be relatively sensitive to the averaging effects of wide antenna beamwidths, which could become a source of ambiguity in aircraft measurements at the lower radar frequencies.

At Low Grazing Angles. At low grazing angles, below mean sea slope angles of about 10°, sea clutter takes on a different character. The sharp target-like clutter peaks known as *sea spikes* begin to appear on A-scope presentations,[1,32,40,41] and the probability distributions assume a different form.[30,42] Figure 15.11*a* and *b* show the presence of sea spikes in 125-second time histories of returns from a fixed spot, measured off the coast of Florida with a variable-resolution X-band radar looking into moderate (*a*) and calm (*b*) seas at a 1.5° grazing angle.[24] Notice that the *appearance* of the sea spikes is very similar for both moderate and weak wind conditions, although the *amplitudes* differ by almost 40 dB, and the vertically polarized returns appear to be somewhat broader, while the horizontally polarized returns are more spiky, particularly for short pulses in calm seas. These are all characteristics of sea clutter at low grazing angles.

The probability distributions of low grazing angle sea clutter change with wind speed. Examples may be found in the measurements by Trizna of low-angle sea clutter using high-resolution (40-ns) shipboard radar in both the Atlantic and the Pacific oceans.[42] The probability distributions of the clutter cross sections were plotted in the manner shown in Figure 15.12, which shows the distributions of horizontally polarized X-band data at a 3° grazing angle for low, medium, and high wind speeds (in order from left to right). The low-wind trace corresponds to a Rayleigh distribution; the other segmented traces are two-parameter Weibull distributions defined by different parameter pairs. It is clear that the behavior is different and considerably more complex than that shown in Figure 15.2 for higher grazing angles and wider pulses. From the character of the data, Trizna interprets these three-segment traces as showing primarily receiver noise in the lowest branch, distributed (spatially homogeneous) clutter in the middle branch, and genuine sea spikes in the branch containing the highest cross sections, some of which exceed 1000 m^2. For the higher wind speeds and fully developed seas encountered in the North Atlantic, the percentage of sea spikes in this population was found to grow as the 3.5th power of the wind speed, which, interestingly, is the same wind-speed dependence shown by the percentage of whitecaps seen on the surface,[22] which, as noted in Section 15.2, are tracers for breaking waves.

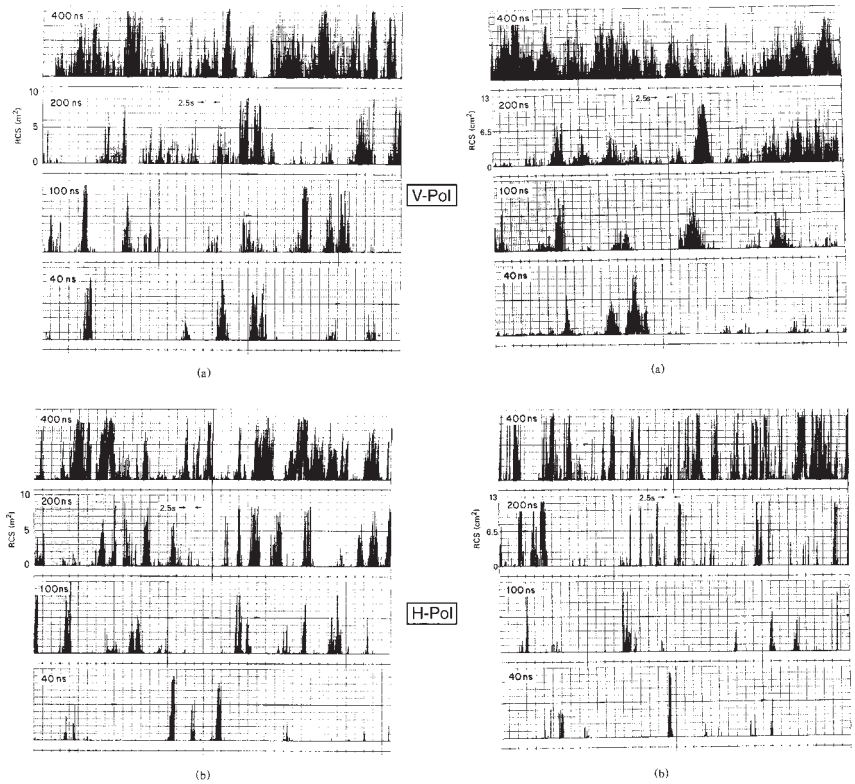


FIGURE 15.11 Sea spikes at X band, 1.5° grazing angle, and various pulse widths: (*a*) sea state 3 and (*b*) sea state 1. Note equal amplitudes at the two polarizations and a 40 dB difference in clutter strength between moderate and weak winds. (*from J.P. Hansen and V. F. Cavaleri*[41])

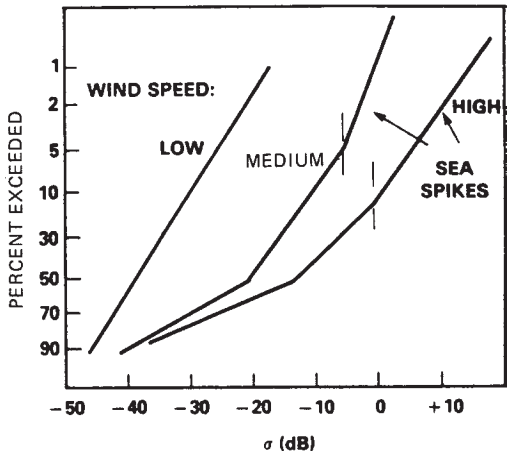


FIGURE 15.12 Examples of clutter probability distributions at low grazing angles (*after D. Trizna*[42])

In comparing statistical results, it should be kept in mind that to the extent that the sea surface may be viewed as a stationary homogeneous process, as it generally is over the duration and spatial extent of any particular experimental event, the scattering cross section may be said to be *ergodic*, which means that the statistical results obtained by time averaging from a small cell are equivalent to an ensemble average from a larger cell, provided that the number of "samples" is the same in the two cases.[43] For this reason, the statistical implications of experimental data can be properly compared only if the details of the sampling procedure are specified. However, the number of samples in most of the experimental results shown thus far have been sufficiently large that the differences between, for example, Figures 15.2 and 15.12, may be considered real and related to differences in grazing angle rather than in resolution cell size. In fact, distributions closely resembling those in Figure 15.12 were obtained much earlier from similar measurements with considerably broader pulse widths.[44] Other measurements have continued to confirm the differences that emerge in A-scope appearance and statistical description of the clutter in this regime of low grazing angles.[45] Some attempts to describe the physical origin of these phenomena will be discussed in Section 15.4 below.

At Very Low Grazing Angles. There is some evidence that sea clutter might drop off more sharply below a *critical angle* in the neighborhood of a degree or so.[4] This critical angle, or *critical range* for a radar at a fixed height, has been observed from time to time since first noted in early observations of sea clutter.[1] The critical angle has been ascribed to interference between direct and (perfectly) reflected rays at the scattering *targets* responsible for the clutter signal, although these targets remain unspecified.[46] Although this simple picture could produce the R^{-7} decay that is sometimes observed, a critical angle often fails to materialize, and when it does, it need not show an R^{-7} decrease with range (equivalent to a fourth-power dependence on grazing angle).[1] An alternative explanation for this behavior, applicable at the higher microwave frequencies, has been suggested based on a *threshold-shadowing* model for upwind and downwind directions.[12,47] This model implies a sharp decrease in the average cross section for grazing angles below a few degrees. In crosswind directions with the radar looking along the troughs of the major waves, a much milder shadowing function will apply, so there should be a clear distinction between the upwind-downwind and crosswind behavior of sea clutter at very low grazing angles.

Examples of clutter behavior at these very low angles may be found in independent measurements at relatively high wind speeds by Hunter and Senior off the south coast of England[48] and by Sittrop off the west coast of Norway.[49] Their results for orthogonal directions relative to the wind are shown in Figure 15.13, along with the predictions of a conventional shadowing function[50] and the threshold-shadowing function.[47] It would appear that a combination of conventional shadowing (which goes as the first power of the grazing angle) across the wind, and threshold shadowing in upwind and downwind directions, can account for the observed behavior of this very low-angle clutter quite well. The decay law for very low-angle clutter should, therefore, depend on the viewing angle relative to the wind direction, so it might occur with powers between the first and the fourth. This is just what has been observed.[51] It should be remarked, however, that *shadowing* at low grazing angles is a complex phenomenon (see below), and the physical origin or even the existence of a critical angle is still open to question. Moreover, there is relatively little good data on very low-angle clutter for other than X-band frequencies, so the general behavior of sea clutter in this angular regime remains uncertain.

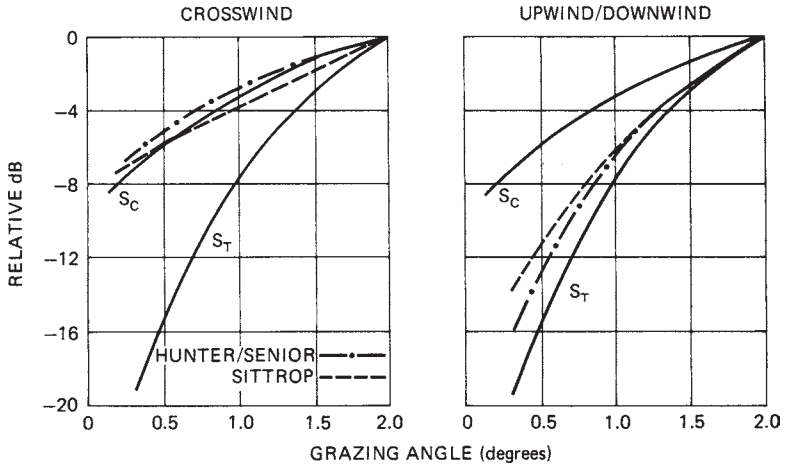


FIGURE 15.13 Differential behavior of very low-angle clutter for orthogonal wind directions: S_c is a conventional shadowing function;[50] S_T is a threshold-shadowing function.[47] (*after data from I. M. Hunter and Senior*[48] *© IEEE 1966 and H. Sittrop*[49])

At HF and Millimeter-Wave Frequencies. All the measurements described above were made at microwave frequencies between UHF (428 MHz) and K_a band (35 GHz). High-frequency (HF) radars usually operate in the frequency range between about 5 and 30 MHz, corresponding to wavelengths between 60 and 10 m, respectively. Since the operation of such radars takes place either by the ground wave or over ionospheric (*sky-wave*) paths spanning great ranges, the grazing angles tend to be small (between 0° and 20°). For these wavelengths and grazing angles, measurements by Crombie indicate that the scattering from the sea surface was the result of scattering from sea waves of one-half the radar wavelength,[5] i.e., "Bragg" scatter. In the years since these early measurements, there has been considerable activity in the field of HF radar and HF clutter,[52,53] and the results can be summarized as follows: For vertical polarization, the major energy of the HF clutter signal appears in spectral lines displaced to either side of the carrier frequency by the frequency of sea waves having a wavelength equal to half the HF wavelength λ (in meters). The relative strengths of the plus and minus lines are determined by the proportion of advancing and receding Bragg-resonant wave components in the clutter cell. Provided the wind speed is greater than about $\sqrt{3\lambda}$ kt (with λ in meters) and the sea is fully developed, the clutter cross section σ^0 is about –29 dB and is relatively independent of wind speed and frequency. (The definition of σ^0 in HF radar is complicated by problems in properly defining antenna gains for ground-wave and sky-wave paths and by propagation effects due to the ionosphere.) The clutter spectrum tends to fill in around and between the lines as the wind picks up. For horizontal polarization (which is possible only for sky-wave paths over which the plane of polarization can be rotated by the Earth's magnetic field), the cross section is much smaller and shows the characteristic fourth-power decay with a decreasing grazing angle. For these HF wavelengths of tens of meters, the sea is relatively flat, and the scattering laws are simple. A detailed discussion of HF radar may be found in Chapter 20.

At the other end of the potentially useful radar spectrum, in the millimeter-wave band, the few published measurements of radar clutter lead to the conclusion that millimeter-wave backscatter behaves in much the same manner as backscatter at the lower microwave frequencies. This was suggested by the K-band curves shown in Figure 15.5 for moderate wind speeds and further supported by some older shipboard data at frequencies between 9 and 49 GHz.[54] It should be noted that for maritime radars clutter signal paths lie close to the sea surface, where the atmospheric and water-vapor densities are highest. This means that at these higher frequencies the clutter signal will be strongly affected by atmospheric absorption effects, and consequently the surface-related cross section inferred from the received signal strength in any given measurement will depend upon the path length. Moreover, the role of sea *spray* in both scattering and absorption will certainly be more important than at the lower microwave frequencies.

It is difficult to find clutter data at frequencies above K_a band, although *H*- and *V*-polarized returns at 95 GHz at a grazing angle of 1° were reported, both with values close to –40 dB.[55,56]

The Spectrum of Sea Clutter. The scattering features producing sea clutter are associated with a surface subject to several types of motion. The features themselves may be moving with small group or phase velocities over this surface while the surface itself is moved, in turn, by the orbital velocities of the larger waves passing across it, or the feature may be advected at the velocity of the wave system supporting it. The scatterers might even be detached from the underlying surface, as in the plumes emitted at the crests of breaking waves, and move at speeds greater than the wave system itself.[57] At higher radar frequencies and in strong winds, the possibility of scattering from spray carried by the wind field above the surface must be considered. All of this complex motion shows up in a doppler shift imparted to the scattered electromagnetic wave. Unfortunately, there is as yet little detailed physical understanding of the complicated phenomenology of sea clutter spectra.

Measurements of microwave clutter spectra for real seas have been reported in the literature for aircraft measurements of the spectral shape alone,[58,59] fixed-site shore measurements showing a shift in the spectral peak,[60,61] and measurements from ships at intermediate grazing angles.[45] Other measurements of sea clutter spectra include those made at much lower frequencies in the HF band, as described in the last section; those made under artificial conditions in wave tanks,[62] whose application to real-sea conditions is uncertain; and those from other fixed-sites at high resolution and short averaging times, to be discussed later.

As it turns out, microwave sea clutter spectra have a rather simple form at the lower grazing angles. Figure 15.14 illustrates typical spectral behavior at the two polarizations, based on data collected by Pidgeon for C-band clutter looking upwind at a few degrees grazing.[60]

The peak frequency of the upwind spectrum appears to be determined by the peak orbital velocity of the largest sea waves, plus a wind-dependent velocity increment containing, but not entirely explained by, wind-induced surface currents. This *peak orbital velocity* is taken to be that of the major waves and may be obtained in terms of the significant height $H_{1/3}$ and peak period T' (Section 15.2) from the expression

$$V_{\text{orb}} = \pi H_{1/3} / T' = 0.15\ U \text{ (m/s)} \tag{15.10}$$

The approximate dependence on wind speed U was found by substituting $H_{1/3} = 6\ h_{\text{rms}}$ from Eq. 15.6, assuming a *fully developed sea*, and T' from Eq. 15.5. To this, there must

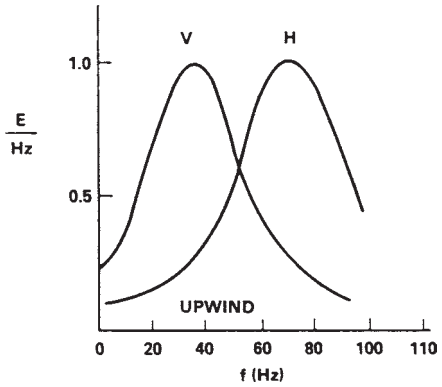


FIGURE 15.14 Qualitative behavior of doppler spectra of sea clutter looking upwind at low grazing angles (*after C-band measurements by V. W. Pidgeon*[60] *© American Geophysical Union 1968*)

be added a wind-drift velocity of about 3% of U and a fixed *scatterer* velocity (which appears to be about 0.25 m/s in the X- and C-band measurements[60,61,63]). Summing these components yields the virtual doppler velocity at the peak of the clutter spectrum for the particular case of a *vertically polarized X- or C-band* radar looking upwind at low grazing angles:

$$V_{\text{vir}} = 0.25 + 0.18\ U\ \text{(m/s)} \tag{15.11}$$

(As noted earlier, care must be taken whenever wind speed is used to parameterize a process that depends on waveheight. There is an unambiguous relation only for a fully developed sea in the absence of swell.) The remaining properties of the clutter spectrum can now be discussed in terms of V_{orb} and V_{vir}. For example, the spectral peak for *horizontal* polarization follows a similar linear dependence on U, only with a coefficient lying somewhere between 0.20 and 0.30, as may be noted in the sketch shown in Figure 15.14. The reasons for the differences between the spectra for the two polarizations are not as yet clear, although the tendency of the H-polarization spectra to lie at a higher frequency is likely due to the preferential source of H-polarization returns in faster-moving wave structures.[45,64]

The (half-power) *width* Δ of the clutter *velocity* spectrum is quite variable, depending on such things as radar polarization and sea conditions. It seems most closely related to the peak orbital velocity given by Eq. 15.10. Nathanson shows a plot containing spectral widths at both polarizations from several investigators over a wide assortment of unspecified sea conditions.[65] The points are widely scattered, but the dependence on wind speed is given roughly, with a rather large variance, by the expression $\Delta \sim 0.24\ U$ (m/s), which is just the orbital velocity in Eq. 15.10 with a coefficient about halfway between the values for vertical (~0.15) and horizontal (~0.30) polarization. For look directions away from upwind, the peak doppler follows a cosine dependence quite closely, going to zero at crosswind aspects and turning negative downwind.[29] The *width* of the spectrum appears to remain relatively constant.

The details of the clutter spectrum show little dependence on either the radar frequency or the grazing angle, at least for angles less than about 10°. In reviewing the results of measurements at four frequencies—UHF, L, C, and X bands—Valenzuela and Laing[59] noted a relatively weak tendency of clutter bandwidth to decrease with increases in frequency between the UHF and X bands and grazing angles between 5° and 30°. Since both of these variations can likely be accompanied by a change in the size of the radar footprint on the surface, they might be due to a dependence on resolution cell dimensions, although the other workers found that the pulse length had little effect on clutter bandwidth for values between about 0.25 and 10 μsec.

Spectra obtained with *short* averaging times disclose something of the origins of the clutter spectrum. Figure 15.15 is a sequence of 0.2 second spectra obtained by Keller et al.[66] with a coherent vertically polarized X-band radar operating at a grazing angle of 35° and a resolution cell size of about 10 m^2. The zero-doppler reference in this figure was located arbitrarily at –116 Hz. The spread along each line

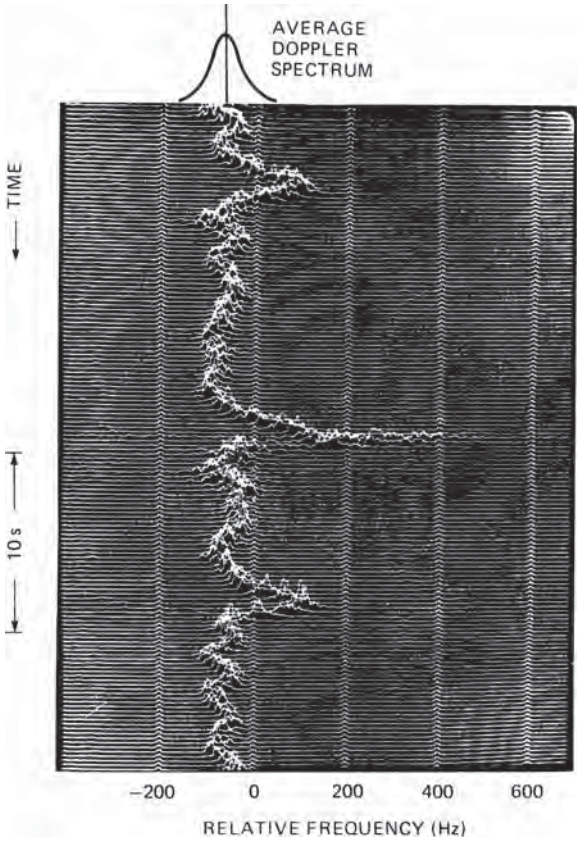


FIGURE 15.15 Short-time averaged doppler spectra at X band for an intermediate grazing angle of 35°; spectra computed at 0.2 sec intervals (shallow-water data is from a fixed site on a pier.) (*from W. C. Keller et al.*[66])

is due to the small-scale wave motions on the surface, while the larger meanders are induced by the velocities of large waves moving through the measurement cell. The wind speed was about 8 m/s, and a doppler shift of 100 Hz corresponds to a radial velocity of 1.6 m/s. The average clutter spectrum expected for this wind speed and grazing angle, with bandwidth estimated from Eq. 15.10, is included in the sketch shown in Figure 15.15. The large spectral spike appearing in the center of the display is no doubt due to a wave breaking in or close to the measurement cell. The doppler velocity for this spike suggests a peak scatterer velocity about equal to the wind speed, which would correspond to the velocity of the longest waves on the surface. Although such events are relatively rare in a fixed area of 10 m^2, they should occur quite frequently within a large surveillance cell and might often have large scattering cross sections associated with them. Similar records may be found in reference Ward et al.[29]

Other Environmental Effects.

Rain. Early evidence of the effect of rain on sea clutter was mainly anecdotal; for example, radar operators would report that sea clutter tends to decrease when it starts to rain. However, there has been little in the way of reliable, quantitative experimental information about the interaction between rain and wind-driven sea clutter in the open ocean. Laboratory measurements by Moore et al.[67] with artificial "rain" suggested that for light winds the backscatter level increased with the rain rate, while for heavy winds rain made little difference. Extensive measurements at K_u band in the open ocean tended to confirm this behavior.[68]

In measurements in natural rain over Chesapeake Bay, Hansen[69] found that even a light rain (2 mm/h) changes the spectral character of sea clutter at moderate wind speeds (6 m/s) by introducing a significant high-frequency component. He also found some evidence in support of the radar operators, at least for the low grazing angles and horizontal polarizations with which most shipboard radars operate. Figure 15.16 compares the correlation function of sea clutter (X band, low grazing angle, H polarization) with and without rain for a 15-kt wind speed and a rain rate of 4 mm/h. The sharp decrease in correlation time in the presence of rain reflects the broadening of the clutter spectrum, although generally there appears to be little quantitative information about the effect of rain on the spectrum of sea clutter.

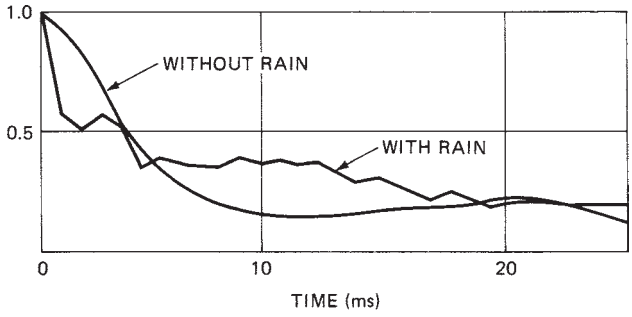


FIGURE 15.16 Effect of rain on the correlation function of wind-driven sea clutter: X band, horizontal polarization, wind speed 15 kt, rain rate 4 mm/h (*from J. P. Hansen*[69])

The production of sea clutter by rain falling on a "calm" surface in the absence of wind was also investigated by Hansen, with the results shown in Figure 15.17.[69] A high-resolution X-band radar (40-ns pulse, 1° beamwidth), operating at a grazing angle of about 3°, viewed the backscatter from a fixed spot on the windless surface of Chesapeake Bay as the rain steadily increased from 0 to 6 mm/h. The cross sections for vertical and horizontal polarizations were quite different for low rain rates but tended to merge at a rain rate of about 6 mm/h. The magnitude of this *splash* cross section rose to a σ^0 of about –40 dB, corresponding to highly averaged wind-induced cross sections at this grazing angle for winds of about 10 kt. Further laboratory[70] and theoretical[71] studies have shown that the major scattering feature under these conditions is the vertical *stalk* that emerges shortly after drop impact. Moreover, these studies suggest that the *V*-polarized returns from raindrop splashes should be only mildly sensitive to the rain rate, while the *H*-polarized returns should show a strong dependence on both the rain rate and the drop-size distribution. Something of this behavior may be seen in the data in Figure 15.17. Open ocean measurements at K_u band and at much higher grazing angles[72] show sufficient variability with wind speed, rain rate, and grazing angle to leave the uncertainties that opened this section largely unresolved.

In addition to scattering from the raindrop impacts, the distribution of raindrops in the volume of the atmosphere above the surface can have two additional effects on sea clutter—as an absorber/scatterer over the radar propagation path, which is well-understood, and as a mass-additive to the wind, affecting momentum transfer to the surface and thus the excitation of wind waves themselves, which is less well-understood.[73]

Atmospheric Ducting. Another topic in sea clutter that has been little explored is the role played by *propagation effects* within the atmospheric boundary layer lying over the sea surface. The effects of atmospheric absorption have been noted above in connection with millimeter-wave clutter, but at very low grazing angles, the ray paths joining the radar to the surface become very sensitive to refractive inhomogeneities in the atmospheric boundary layer. Over distances approaching and

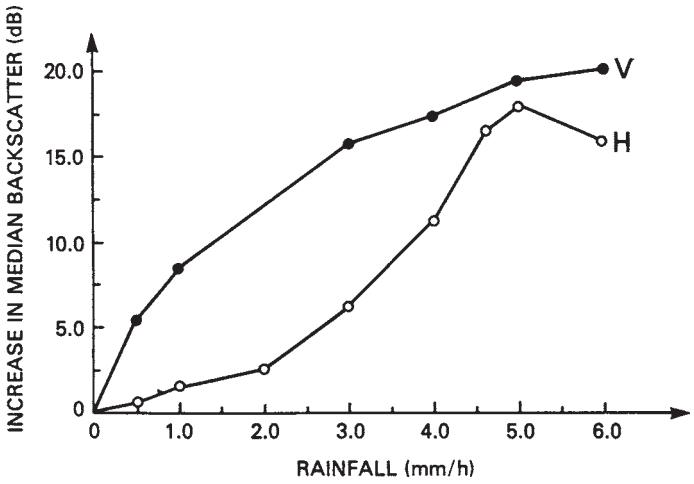


FIGURE 15.17 Sea clutter produced by rain splashes alone on a calm surface (20 dB corresponds to about $\sigma^0 = -40$ dB) (*from Hansen*[69])

beyond the conventional optical horizon, such perturbations could produce strong focus-defocus variations along the surface illumination profile[74,75] or a general rise in the local grazing angle.[47] Figure 15.18 gives an experimental example of the effect of *ducting* on very low-angle sea clutter.[51] The grazing angle given as the abscissa is actually a plot of inverse range, so the lifting of the cross section by ducting over an order-of-magnitude span of ranges is very likely due to a rise in the local grazing angle produced by refraction in the evaporative layer (first 10 meters or so above the surface).[47] Such effects should be suspected whenever the radar propagation path extends beyond the *optical* horizon.

Shadowing. The possibility of shadowing must be considered seriously whenever the sea is viewed at grazing angles smaller than the rms slope angle of the sea surface. Some examples were discussed earlier in connection with the behavior of sea clutter at low grazing angles in Figure 15.13. In fact, the sharp falloff of the *nonducting* data in Figure 15.18 might be further evidence of the *threshold shadowing* mentioned earlier. However, the common idea of shadowing derives from the geometrical optics concept of a sharp transition between light and darkness. By considering the implications of *diffraction* at the wave peaks, it is possible to determine the domain of radar frequencies and wind speeds over which the concepts of geometrical optics may be applied. This was done in Wetzel,[12] where it is shown how diffraction, rather than geometrical shadowing, controls propagation into and out of the troughs of the waves under many of the usual frequencies and wind speeds encountered in practical radar operations at low grazing angles. For example, shadowing will take place at K_a band for any winds above 15 kt, yet will hardly ever occur at L-band frequencies. Later analytical solutions based on numerical methods explored and expanded the idea of shadowing over the sea surface under highly idealized,[76] and more general[77] conditions.

Surface Currents. The most obvious effect of a current on sea clutter would be a shift in the peak of the doppler spectrum, analogous to the contribution of the 3% wind-drift current mentioned in connection with Eq. 15.11. Another effect is related to

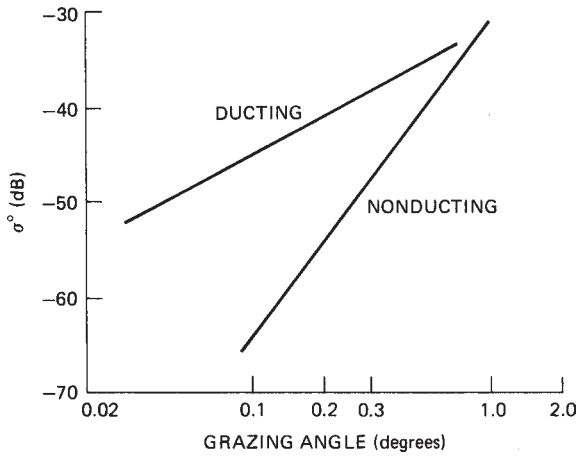


FIG.URE 15.18 Effect of ducting on low-angle clutter; wind speed about 10 kt (*after F. B. Dyer and N. C. Currie[51] © IEEE 1974*)

the fact that the excitation of the surface-wave system depends on the local *apparent wind* at the surface, so there can be significant differences in waveheight according to whether the wind is blowing with or against the current. According to Eq. 15.6, waveheight is proportional to the square of the wind speed, so in the Gulf Stream, for example, with a current of 4 kt flowing north, a 15-kt northerly blowing against the current will raise a sea three times as high as a 15-kt southerly blowing with the current. Even with no wind, the presence of strong *current shears* can produce highly agitated surfaces. Shipboard observers have reported bands of roaring breakers passing by on an otherwise-smooth surface, presumably produced by powerful surface-current shears associated with large-amplitude internal waves.[78] In a more subtle way, modulated currents are held responsible for synthetic aperture radar (SAR) images that contain the expression of bottom topography produced by the Bernoulli Effect in shallow waters.[79] In each of the examples cited above, the current produces a change in the surface roughness, which can be expected to give rise to a change in sea clutter cross section.

Contaminants. The idea of pouring oil on troubled waters is a familiar one: the angry surface will smooth and subside. In another age, the survival-gear locker of every sailing ship would contain a bottle of oil to quiet the sea in a storm. Although the effectiveness of this procedure has always been somewhat controversial, there is no question that oil can produce a *slick* of smooth water at relatively low wind speeds. In fact, biological oils, produced by bacteria, algae, and plankton, can be found everywhere on the world's oceans and form natural slicks in those regions that combine the greatest oil concentration with the lowest wind speeds, e.g., close to continental shorelines.[80] Man-made contaminants can, of course, have the same effect. A layer of oil only one molecule thick will significantly affect the ability of the surface to support wave motions, but this layer must be continuous. The adjacent molecules then sense each other and form a film that is resistant to horizontal compression. The surface elasticity is changed, a type of longitudinal viscosity is introduced, and the surface becomes stabilized against the growth of short waves up to several inches in length.[81,82]

To the extent that radar sea clutter is produced by small-scale surface roughness (at grazing angles less than about 80°), the presence of oil on the surface should lead to a measurable decrease in clutter cross section. But, as noted above, the reduction of small wave motions requires the existence of a *continuous* monolayer; slick formation is a go–no-go process, and so slicks will tend to have relatively sharp boundaries. In operating the NRL 4FR system as a synthetic aperture radar to obtain images of the slicks produced by oil spills, Guinard found that the slicks were well defined, that it took very little oil to maintain a visible slick, that vertical polarization provided much greater contrast than did horizontal, and that the slicks were quenched by winds and currents.[83] Although signal strength was not recorded in this imaging experiment, later measurements at X and L bands by others[84] indicated that at the higher grazing angles (about 45°) the clutter reduction produced by the types of oil occurring in *natural* slicks was rather small, on the order of a percent. Since slicks are dispersed by the wind and associated wave action at wind speeds greater than about 10 kt, the effect of natural slicks on clutter may not be clear because they tend to occur in the regime of low wind speeds where the sea surface is already ill defined.

The celebrated *sun glitter* measurements by Cox and Munk[85] gave a quantitative measure of the effect of contaminants on the surface slopes in open water, showing that the wind-generated component of the rms slope of "oiled" waters is significantly

smaller than that of "clean" water. The heavy commercial oils used in their experiment were effective in suppressing small-scale waves over a range of wind speeds well beyond those which would normally disperse the lighter natural oils, so the effect of oil spills on sea clutter should be expected to extend to the higher wind speeds. In fact, at these higher wind speeds, the depression of radar backscatter by such oils at X and K_a bands can reach 10 to 20 dB at intermediate grazing angles between 30° and 60°.[86,87]

15.4 THEORIES AND MODELS OF SEA CLUTTER

In addition to providing an intellectual basis for "understanding" sea clutter phenomena, a complete theory of sea clutter should ideally provide accurate *a priori* predictions of all aspects of clutter behavior under all possible environmental conditions. In spite of over 60 years of effort, the theory of sea clutter does neither of these tasks very well, as we will see in this section.

In the development of models of sea scatter based on physical theory, there are essentially two basic and distinct approaches. Historically, the first approach assumed sea scatter to have its origin in scattering *features*, or obstacles, actually present on or near the sea surface. Early scattering models included rain (to model spray),[88] smooth circular metallic disks,[46,89] arrays of semi-infinite planes,[90] and fields of hemispherical bosses,[91] to name a few. Obviously, the choice of these scattering obstacles related more to the preexistence of convenient scattering solutions for these shapes than to insights gained from observing the sea. Since then, feature models have sought greater reality by considering wedge shapes, as suggested by the sharp crests of Stokes waves observed on most natural water surfaces[12,32,92,93] and the shocks and plumes suggested by the properties of wave groups and the hydrodynamics of breaking waves.[12,57]

The other approach to theoretical modeling derives the scattered field from a global boundary-value problem (GBVP) in which the sea as a whole is considered a boundary surface whose corrugations are described by some kind of statistical process. An enormous literature is devoted to the theory of surface scatter from this point of view, stemming from the importance not only of radar sea scatter, but also radar ground scatter and sonar *reverberation* (the acoustic equivalent of radar clutter) from both the surface and the bottom of the sea. Because the GBVP approach leads to the analytical expression of the Bragg scattering hypothesis that has dominated the theory of sea scatter since the late 1960s, a brief explanation of some of the central ideas is included below.

Theories Based on Global Boundary-Value Problems. General formulations of the GBVP, though elegant, are of little practical value, and some kind of approximation is necessary to obtain useful quantitative results from them. The methods of approximation relate to the two methods of formulating the GBVP:

1. Small-amplitude approximations (sea waveheights much smaller than the radar wavelength) are used *ab initio* with Rayleigh's hypothesis, in which the surface boundary conditions are employed to match an angular spectrum of outgoing plane waves to the incident field.[94–96]
2. A general integral formulation based on Green's theorem is pursued either in a small-amplitude approximation[6,97] or under the assumptions of physical optics (surface curvatures much greater than the radar wavelength).[98–100]

In formulation 1, sometimes called the *small-perturbation method (SPM)* and associated most often with the work of Rice,[94] the surface displacements are assumed *everywhere* to be much smaller than the radar wavelength, so the method is directly applicable only to such cases as HF scattering with wavelengths of tens of meters, at low to intermediate wind speeds, and with waveheights of a few meters at most. The solution is in the form of a power series in the ratio of sea waveheight to radar wavelength, and it predicts the first-order Bragg lines and second-order spectral filling around the lines that were mentioned in the earlier section on HF sea clutter.

On the other hand, the various integral formulations referenced above usually begin with a very general expression for the fields scattered from the sea surface, which are squared and ensemble-averaged over realizations of the sea surface to provide the average power returned to the radar antenna, and then normalized to the illuminated area as in Eq. 15.7. Despite the generality of the initial formulations, most of the final expressions for σ^0 either appear in, or can be put into, a form represented schematically by the following simplified one-dimensional expression (see Holiday et al.,[10] Bechmann and Spizzichino,[99] Fung and Pan,[101] and Valenzuela,[102] for example):

$$\sigma^0(\psi) = Ak^2 F_p(\psi) \int_{-\infty}^{\infty} dy\, e^{i2k_1 y} [e^{-4k_2^2 h^2 [1-C(y)]} - e^{-4k_2^2 h^2}] \tag{15.12}$$

where A is a constant; $k_1 = k \cos\psi$ and $k_2 = k \sin\psi$ where k is the radar wavenumber $(2\pi/\lambda)$; $F_p(\psi)$ is a function of polarization p, grazing angle ψ, and the electrical properties of seawater; h is the rms sea waveheight; and $C(y)$ is the surface correlation coefficient. Of course, the reduction of a complicated boundary-value problem to so simple a form requires assumptions about both the surface fields and the distribution of the sea heights (which is gaussian to a good approximation[20]). But while the SPM approach mentioned above requires the ratios h/λ to be small right at the start, GBVP theories derived from expressions resembling Eq. 15.12 have no *a priori* restrictions on surface heights.

The statistical properties of the sea surface enter through the correlation coefficient $C(y)$ appearing under the integral sign in the exponential in the brackets, and by expanding this exponential in Eq. 15.12, it may be written

$$\sigma^0(\psi) = Ak^2 F_p(\psi)\, e^{-4k_2^2 h^2} \sum_{n=1}^{\infty} \frac{\left(4k_2^2\right)^n}{n!} W^{(n)}(2k_1) \tag{15.13}$$

where

$$W^{(n)}(2k_1) = \int_{-\infty}^{\infty} d\tau\, e^{i2k_1\tau} [h^2 C(\tau)]^n \tag{15.14}$$

Small-amplitude Approximation. In the limit of small ratios of rms waveheight to radar wavelength or, more specifically,

$$2kh \ll 1 \tag{15.15}$$

only the first term in the series in Eq. 15.13 survives, and the cross section assumes the very simple form

$$\sigma^0(\psi) = 4\pi k^4 F_p'(\psi) W^{(1)}(2k\cos\psi) \tag{15.16}$$

where the constant A has been made explicit and F_p has absorbed a $\sin^2$ term from the series. $W^{(1)}$ is the Fourier transform of the surface correlation function, which makes it the sea wavenumber spectrum (discussed in Section 15.2) evaluated at twice the (surface-projected) radar wavenumber, which defines a *Bragg (or half-wavelength) resonance*. Except possibly for the details of the angle factor F_p, Eq. 15.16 is equivalent to the result obtained by the SPM discussed above, and although it is sometimes felt that its derivation from a surface integral provides some potential for greater generalization, it carries with it all of the same restrictions.

Before proceeding further, it is instructive to look a little more closely at the implications of the mathematical expressions in these formulations. Note that in Eq. 15.12 the expression in the brackets under the integral is the only place at which the surface-wave properties of the sea appear. That is, the cross section is simply proportional to the Fourier transform of a sea-surface functional, so the radar acts as a filter tuned to the "spatial frequency" $2k\cos\psi$, extracting that line from the spectrum of whatever assortment of scatterers the sea surface functional expresses, whether they be long swell components, short-wavelength noise, localized scattering features, or a chaotic tumble of water balls. Only under quite special circumstances will there actually be an identifiable surface-wave at that "frequency" that would justify the term *Bragg resonance* in its original sense, which, after all, was a resonance in an ordered crystal lattice of discrete scatterers. Although authors often refer to "free Bragg waves," such objects are found primarily among the *parasitic capillaries* referred to in Section 15.2 or among the *ring waves* propagating out from the impact of a falling drop. This raises a question about the meaning of such often-used terms as *Bragg wavelets, Bragg patches*, and so on, as if such things had a real existence outside of their emergence as an artifact of a filter operation. This confusion of concepts might be avoided by visualizing the radar as extracting the "Bragg line" from the spectral compositions of those surface features that contain a wave of length $\lambda/2\cos\psi$ in their Fourier *representation*, independent of the existence of such a wave as a real object.

Nevertheless, this direct linear relation between radar cross section and the oceanographers' descriptor of the sea surface has had a powerful influence on thinking about the physical origins of sea clutter. It is appealing in its simplicity, and it suggests a direct way both to predict radar clutter from measurements or forecasts of the sea spectrum and, inversely, to use radar backscatter measurements to provide remote sensing of the sea surface for oceanographic and meteorological applications—provided, of course, that it correctly describes this relationship.

Although Eq. 15.16 successfully predicts (to 1st-order) the clutter returns at HF frequencies, at *microwave* frequencies the small-waveheight assumption on which this model rests is violated on any real sea surface. The small-waveheight condition expressed by Eq. 15.15 means that at X band, for example, the maximum departure of the sea surface from a flat plane must be much smaller than 3 mm.

Other Calculational Strategies. Instead of expanding the exponential in the integrand of Eq. 15.12, it should be possible, at least in principle, to replace $C(y)$ directly by the Fourier transform of $W(K)$ (the inverse of Eq. 15.14 for $n = 1$), thus providing a direct functional relationship between the radar cross section and the sea wave spectrum without the restrictions of a small-amplitude approximation. This cumbersome approach involves extensive computations even to obtain limited results in individual cases, as shown in work by Holliday et al.[10]

In another limiting case, the basic integral formulation of the GBVP is solved in the optical approximation (large k), resulting in an expression commonly called the

specular return, because its origin may be traced to pieces of the surface that provide a reflection point for the incident wave.[99,102] This expression is written for a gaussian sea surface in the form

$$\sigma^0(\psi) = (|R|^2/s^2)\csc^4\psi \exp[-\cot^2\psi/s^2] \tag{15.17}$$

where s is the rms surface slope and R is the flat-surface reflection coefficient for normal incidence. This is the type of scattering alluded to in connection with the high grazing angle returns discussed in Section 15.3; the tendency of σ^0 to level off for grazing angles close to 90° (see Figures 15.3 and 15.4) may be ascribed to this mechanism.

From what has been said thus far, it can be seen that strict analytical solutions via the GBVP approach appear to run into dead ends: intractable formal expressions in the form of Eq. 15.12, small-amplitude approximations in the form of Eq. 15.16 that make little sense for microwave scattering from real sea surfaces, or optical limits such as Eq. 15.17 that relate to the probability densities of specularly reflecting surface slopes. It appears, therefore, that the use of integral formulations in the practical solution of the sea clutter problem at microwave frequencies will require something more.

Composite-Surface Hypotheses. Since it is not clear how to extend straightforward GBVP solutions beyond the limiting approximations described above, a heuristic model was developed that viewed the sea as a carpet of Bragg scattering "wavelets" modulated by the motions of the larger waves on the surface.[102–104] This *composite surface model* is often referred to as the *two-scale model*, in which it is imagined that the surface-wave spectrum can somehow be separated into two parts, one containing low-amplitude "Bragg scattering wavelets" whose integrated rms waveheight satisfies the conditions of Eq. 15.16 and another that contains only the longer waves that tilt and stretch and otherwise modulate the Bragg waves, affecting the Bragg scatterers through a *modulation transfer function*,[105] as well as providing a specular component resembling Eq. 15.17. Other assumptions include that (1) the correlation lengths of the short Bragg waves be long enough that a resonant interaction is possible, but short enough that adjacent areas on the surface contribute to the total signal in random phase (note how "Bragg waves" are viewed here as physical objects); and (2) the long waves that tilt and modulate the short waves have radii of curvature sufficiently large that the curvature over the correlation length of the "Bragg patches" is small in some sense. In its simplest and most commonly used "tilt" form, it interprets $\sigma^0(\psi)$ in Eq. 15.16) as the cross section of a patch with *local* grazing angle $\psi = \sigma^0 + \alpha$, where α is the local wave slope and ψ_0 is the mean grazing angle. For the simple one-dimensional case, this quantity is averaged over the sea slope angle distribution $p(\alpha)$, yielding

$$\bar{\sigma}^0(\psi) = \int_{-\infty}^{\infty} \sigma^0(\psi_0 + \alpha)\, p(\alpha)\, d\alpha \tag{15.18}$$

For a more general two-dimensional sea, the local grazing angle is a function of the slopes in and normal to the plane of incidence, so for each polarization p, the angle function $F_p(\psi)$ in $\sigma^0(\psi)$ becomes a complex mixture of the angle functions of both polarizations. Plant[104] provides a comprehensive discussion of this model. Extensions of this model into what might be called a *three-scale model* are guided by the same ideas that led to the *two-scale model*.[72,106] An additional ad hoc spectral partition is introduced between the longest and shortest wave components of the sea spectrum, but this produced only modest improvement.[106]

Although such composite-surface models may leave the impression that they have emerged as a rigorous product of an integral formulation of the GBVP, it is clear that they are not really scattering *theories*, but instead scattering *pictures* assembled from a group of more or less plausible assumptions. But with the failure of the more formal GBVP theories to provide a general framework for predicting and understanding sea clutter, these models have become the basis for most analytical approaches to microwave scattering from the sea.

Figure 15.19 compares a sample of NRL 4FR data taken at high wind speeds (> 22 kn) with the predictions of the pure *SPM Bragg* model (in the form of Eq. 15.16) for V-polarization and the *two-scale model* (in the form of Eq. 15.18) for H-polarization. The wave spectrum used was the Phillips spectrum given in Eq. 15.4. Historically, comparisons of this type have been used often to provide support for the Bragg scattering hypothesis,[102,104] and the agreement often looks good, especially for vertical polarization at the higher wind speeds. Yet why this is so remains a puzzling curiosity. In this example, the wind speeds are high so the sea will be rough, but it was noted above that the SPM approximation used in Figure 15.19 requires waveheights much less than a centimeter, so this approximation is totally invalid for these data. Moreover, the Phillips spectrum was used as the sea surface spectrum, but there is no evidence that this spectrum holds down to the required capillary wavelength of about 1.5 cm. In fact, the nature of the sea spectrum in this range remains uncertain and has been referred to as "one of the most exciting unsolved problems of the sea surface."[16] The primary effect of the *two-scale model* is simply to raise the inapplicable SPM values for H-polarization, which are dropping sharply with grazing angle, by including higher local angles via Eq. 15.18. There would be little sensible effect on the more constant vertical returns. Finally, the polarization differences themselves have been shown to derive entirely from the behavior of the reflection coefficients at the underlying surface (see, for example, Wright[6] and Wetzel[12]), and so are not an intrinsic part of the Bragg hypothesis and would apply to *any* small perturbation on the surface. Nevertheless, agreement between measurement and prediction of the type illustrated in Figure 15.19

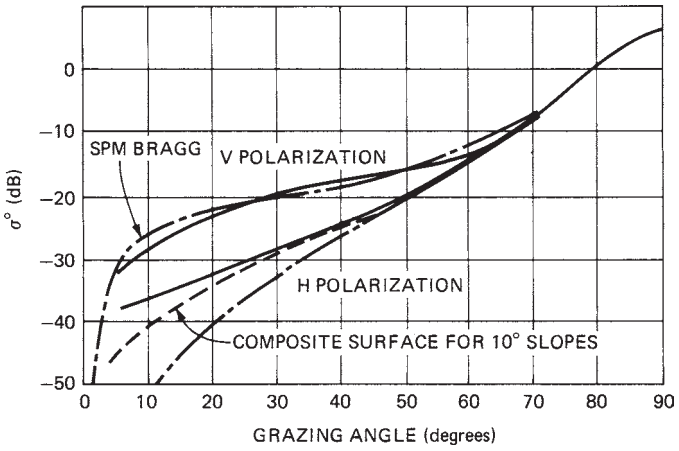


FIGURE 15.19 Comparison of the predictions of the Bragg hypothesis with NRL 4FR data (heavy lines) for higher wind speeds (> 22 kt): dash-dotted lines, SPM Bragg; dashed line, two-scale model for H-polarization only, assuming 10° mean sea slopes (*after Valenzuela*[102])

has kept alive credibility in the Bragg scattering hypothesis in spite of the clear violations of conditions and nonsequiturs noted above, as well as the lack of a proper theory argued from first principles.

The failure of this model to account for sea spikes and other non-Rayleigh returns noted in Section 15.3 has led to an augmentation of the two-scale composite surface to include breaking waves, a presumed source of these returns. One of the latest of these *two-scale-plus* models may be found in Kudryavtsev et al.,[107,108] where breaking-wave effects are incorporated analytically through the Phillips expression[23] for the density of breaking fronts as a function of wind speed (the parameter $\Lambda(c)$ noted in Section 15.2), with the scattering behavior based on Wetzel's plume model (see "Scattering by Surface Features"). The result is a significant improvement in sea scatter predictions, emphasizing the importance of breaking in the scattering scene. An example of this improvement is shown in Figure 15.20, where experimental data for the *polarization ratio* relative to wind direction was obtained from the Polrad96 program.[107] One of the major claims for Bragg models has been their apparent ability to agree loosely with the observed polarization ratios of sea clutter returns, and in Figure 15.20, as in Figure 15.19, the data were compared to the predictions of the SPM (pure) Bragg model and the two-scale model, while adding the *two-scale-plus model*, which incorporates the effects of breaking waves. In this case, the Bragg model and the two-scale model derived from it clearly fail to describe the observed behavior of the polarization ratio, while the inclusion of breaking waves (full model) provides a surprisingly faithful reproduction of the data, even at the high grazing angle of 50°. The X-band data illustrated previously in Figure 15.10*a*,*b* can also be shown to corroborate the predictions of this model for a similar wind speed and grazing angle, thus further supporting the suspicion that breaking-wave events will contribute importantly to sea clutter over *most* of the "rough surface" regime of grazing angles at these wind speeds.

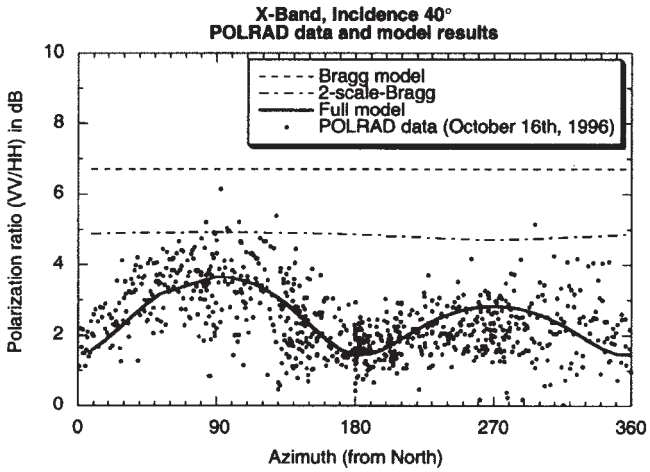


FIGURE 15.20 Polarization ratio versus wind direction for X-band radar at 50° grazing angle and wind about 20 kt. *Points:* experimental data from Polrad96 experiment; *dashed line:* pure SPM Bragg; *dashed-dotted line:* two-scale Bragg; *solid line:* two-scale-plus model including breaking waves (*after Kudryatsev et al.*[107] *© The American Geophysical Union 2003*)

Scattering by Surface Features. A breaking wave, with plumes of water cascading down its face and perhaps a halo of spray above, is only one of the rich assortment of scattering elements appearing on the sea surface—including wedges, cusps, microbreakers, hydraulic shocks, patches of turbulence, and gravity-capillary waves (both wind-driven and parasitic)—any or all of which could contribute to the scattered clutter signal.

For example, the common Stokes wave[20] has a quasi-trochoidal structure that resembles a wedge on the surface, so wedge scatter might describe an important aspect of sea clutter.[11,12,92,93] The scattering model is usually some variant of the familiar *geometrical theory of diffraction* (GTD),[109] which is strictly applicable to the backscatter problem only when the edge of the wedge is normal to the plane of incidence. Nevertheless, the cross section predictions at the lower grazing angles for both polarizations show trends similar to the predictions of the Bragg or composite-surface models.[93]

One major problem with all models based on scattering feature simulations is the lack of reliable information about the shapes, sizes, orientations, speeds, lifetimes and statistics of the features themselves. Thus, although there is often guidance from either observation or theory, the predictions of such models will be based on uncertain assumptions about these crucial parameters. As an example, water surface stability arguments prevent the interior angle of a sharpening wave crest from falling below 120°, which then becomes a convenient measure of the wedge angle in wedge-scattering models. In Figure 15.21, the overall scale of wedge scattering as calculated by the GTD was adjusted to locate the cluster of cross sections at the level of the experimental values. Wedges appear to model the qualitative behavior with both polarizations fairly well at the lower grazing angles.

Figure 15.21 also includes two additional simple scattering models for comparison. Lambert's law, mentioned in connection with Figure 15.9*a*,*b*, expresses the cross section in the form $\sigma^0 = A \sin^2\psi$, where A is the *surface albedo*. Choosing $A = -13$ dB (a reasonable value for microwave frequencies) gives a fairly good match to the vertically polarized returns over a wide range of grazing angles. The *facet* model, expressed

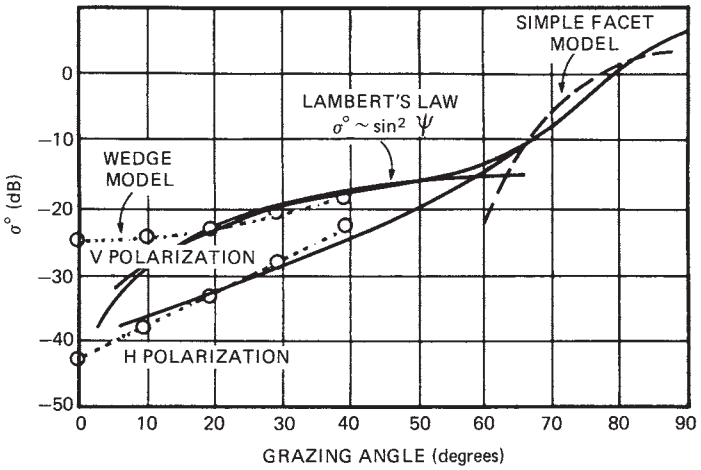


FIGURE 15.21 Comparison of several ad hoc feature models with NRL 4FR data (data the same as shown in Figure 15.19)

by Eq. 15.17 and thought to describe clutter at the higher grazing angles, is shown for 20 kt seas.[102] The general behavior described by these two models seems to agree about as well as any other, although they too must employ arbitrary assumptions to obtain reasonable fits to the data, so the significance of this agreement is difficult to assess. However, one might deduce from Figure 15.21 that sharp things (like wedges) dominate the clutter at small grazing angles, flat things (like facets) at large angles, and generally rough things at the intermediate angles.

The theory of scattering from breaking waves referenced in connection with Figure 15.20 was originally motivated by an attempt to explain the complex behavior of sea spikes at low grazing angles observed by Lewis and Olin.[40] This theory was based on the *plume model* of the most common *spilling* breaker, in which a water plume is emitted at the spilling crest and moves down the front face;[110] the scattering behavior was supplied by multipath illumination involving reflections from the face of the wave.[57] The elaboration of this model explained much of the complex behavior of the observed sea spikes; however, like all other models based on scattering features, it was necessary to make assumptions about the sizes, shapes, and lifetimes of the scattering plumes. These parameters were all inferred from observation of real sea surfaces, and the resulting predictions were surprisingly good. Additionally, its success in the *two-scale-plus* model mentioned above provides further credibility.

Although scattering features have been introduced mainly in connection with low-grazing-angle sea clutter (see Wetzel[12] for a detailed discussion), there is evidence, as noted above, to believe that feature scattering operates at all grazing angles. Considering the failure of scattering theories formulated as a GBVP to provide any predictions beyond those in certain limiting-case approximations and the precarious nature of the logical infrastructure of the Bragg hypothesis in microwave scattering from a natural sea, it is quite possible that further careful consideration of the actual scattering features present on the sea surface will improve our understanding of sea clutter in the future.

Implications of Surface Geometry. The approximations to the GBVP discussed above were all formulated in the frequency domain, but a time-domain model for sea scatter points to a possible general surface-geometric origin for the surface returns.[13,110] The model introduces the idea that the basic scattering elements of a surface are localized at points of *high surface curvature*, as at the sharp tips of small Stokes waves or at the corner of a plume intersecting the front face of a breaking wave. Using a δ-function as a probing pulse, an approximate expression for the scattering cross section was found as a function of surface curvature $\mathcal{C}(\mathbf{r}(t_o))$, where t_o is the round-trip time from the radar to a point $\mathbf{r}(t_o)$ on the surface

$$\sigma(t_o) = B(\psi, s, t_o)\, \mathcal{C}^2(ct_o/2\cos\psi) \tag{15.19}$$

with B a complicated trigonometric function of grazing angle ψ, surface slope s, and t_o. The factor B peaks at points of specular reflection, so Eq. 15.19 expresses the effects of specular glints as well as sharp curvatures. Although the original theory[13] is based on a physical optics formulation for a scalar field (see Sangston[111] for the vector form), some support for its implications can be gained from a plot of the square of surface curvature over a real sea surface. Figure 15.22*a* shows the *surface slope* measured over a short segment of surface in the Gulf Stream,[112] with the corresponding *curvature-squared* profile below. The clutter cross-section defined for this profile

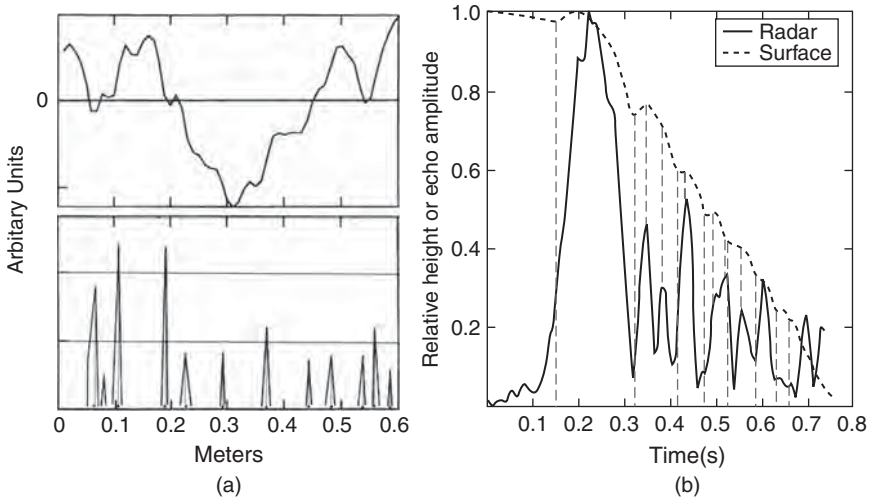


FIGURE 15.22 (*a*) Measured slope (top curve) over a sampled surface in the Gulf Stream,[112] with the corresponding square of the surface curvature below; (*b*) Wave-tank measurement of radar scattering from an evolving breaking wave, correlated with wave surface height variations (*from M. A. Sletten and J. C. West*[114] *© The American Geophysical Union 2003*)

by Eq. 15.19 clearly bears a resemblance to some of the high resolution returns shown in Figure 15.11. Moreover, these spiky returns are correlated with just the kind of surface bumps and wrinkles and slope discontinuities that have been identified as sources of sea spikes in laboratory tank measurements.[11,113,114] An example is shown in Figure 15.22*b*, where an ultra-high-resolution radar looks at the return from a breaking wave whose surface is evolving as plotted above the radar signal. We see that the spikes appear at the points of maximum surface curvature, although the large signal peak is probably a specular reflection from the initial breaking front.

It is often overlooked that the presumed origin of sea scatter is suggested by the kind of measuring instrument being used, which in turn determines the most appropriate theoretical basis—i.e., the "theory of the experiment." If you go to sea with a CW radar—an "averaging wave spectrometer"—you will be selecting a spectral line by long integration times and interpret the origin as a wave effect, namely a "Bragg resonance." On the other hand, if you use a high resolution probe—a radar microscope—the clutter scene will be populated by highly localized scattering events, or *sea spikes*, isolated by a short-pulse, wide-bandwidth signal. In this case, the most appropriate "theory of the experiment" is the time-domain formalism described above.

Other ways of viewing sea clutter in terms of surface geometry characterize the sea and the clutter as fractal processes,[115] or seek parameters describing its complexity by defining a "strange attractor."[116] Unfortunately, these studies do not seem to contain any useful insights into the physical scattering processes at the surface, except, perhaps, to conclude that sea clutter arises from multiple sources, which we already know. On the other hand, the identification of changes in the characteristic *measures* of these processes (e.g., fractal dimension and embedding dimension) have been proposed as a way to identify the presence of targets in clutter.

Numerical Methods. With rapid increases in computer speed, it has become practical to solve certain scattering problems by numerical methods, thereby sidestepping the introduction of the approximations required by the analytical solutions described above. In surface scattering, particularly for the continuous sea surface, the method of choice appears to be some variant of the Method of Moments.[117] An exact integral equation for the *surface currents* excited by an illuminating field is solved numerically over a grid of points, where the flexibility and accuracy of the solution depends, essentially, on the grid spacing, the size of the surface features compared to the illuminating wavelength, the extent of the surface over which the grid is laid, and the efficiency of the computing algorithms used in what are extensive calculations. Once found for an ensemble of surface realizations, these currents can be used in a scattering integral over each of the surfaces to generate an ensemble of *scattered fields*, which are finally averaged into a surface scattering cross section.[118] Although these methods are often viewed as the "gold standard" for doing scattering calculations, their complexity and difficulty generally restrict their application to idealized or laboratory surface structures, where they simply confirm that Maxwell's equations continue to be verified in scattering experiments. Nevertheless, such numerical simulations can be informative in identifying the source of particular scattering events, such as sea spikes, along with their dependence on such radar parameters as grazing angle, polarization, and frequency.[119,120]

Role of Laboratory Studies. The sea surface is a natural system controlled by the laws of hydrodynamics. But so is a bowl of soup or a tank of water or a rushing stream—all of which might share certain behaviors with a sea surface, while presenting a venue much more amenable to the comfortable investigation of scattering phenomenology. While there is the obvious matter of *scale*, one might entertain the notion that what is found to hold in the small compass of a laboratory wave tank might transfer with little alteration to the open ocean. But this certainly cannot be true; the sea surface is structured by large-scale wind systems, impossible to duplicate under laboratory conditions. Therefore, the laboratory is used almost exclusively to study in detail how an electromagnetic wave interacts with some restricted and well-controlled aspect of surface phenomenology that is thought to be involved in the open sea.

There is a vast literature extending over almost 45 years reporting a wide variety of laboratory experiments on microwave scattering from a disturbed water surface. The earliest appear to be careful small-scale experiments in the mid-1960s to confirm the existence of Bragg scatter at centimeter wavelengths,[6] while at the other end of the experimental scale, the chaos of a full-scale breaker was simulated by submerging a hydrofoil across a 7 meter wide circulating water channel driven by 1.5 megawatts of turbine power.[121] But most laboratory experiments are more modest, involving a long wave tank, perhaps a meter or two across, with the wave system produced by either a controlled wind or a programmed wavemaker.

To illustrate a few of the experiments and their results: Kwoh and Lake[11] measured X-band returns from gentle breaking waves in a wave tank and found that specular and curvature scattering appeared to dominate over Bragg scattering for such surfaces. Keller et al.,[122] measured X-band returns and surface spectrum simultaneously in a wave tank and found Bragg-based theories might have credence at intermediate grazing angles and strong winds, but were unable to account for scattering behavior under other conditions. Sletten and West[113,114] made a two-pronged approach to breaking wave scatter, constructing a metallic model of a breaking wave, comparing its

scattering properties with numerical calculations, and then going to the wave tank to verify the scattering behavior for real water waves. Ericson et al.[123] generated a stationary breaker in a small flow tunnel and compared its scattering properties with a specular point calculation. Coakley et al.[121] set up a hydrofoil in a powerful water channel, generating a large breaking wave front that produced radar returns consistent with Lambert's Law, along with the polarization ratios implied by multipath reflections from the undisturbed front face. These investigators, along with many others, have been exploring radar scattering from water surfaces in ways that can be expected to contribute significantly to our growing understanding of sea clutter.

15.5 SUMMARY AND CONCLUSIONS

In the early days of radar, the importance of knowing the sea clutter environment led to many experiments under a variety of conditions. Variations in quality and completeness of ground truth, calibration of the equipment, and the experimenter's experience led to results that often showed considerable inconsistency and suggested clutter behavior that was sometimes more a function of the vagaries of the experiment than of the physics of the clutter. As data of increasingly better quality accumulated, it might have been expected that the behavior of sea clutter would be established with increasing confidence. This has not always been so, leaving our general knowledge about sea clutter to be summarized roughly as follows: For the higher 50% of wind speeds encountered over the world's oceans (greater than about 15 kt), microwave sea clutter at intermediate to high grazing angles has little dependence on frequency, and the effects of wind speed are uncertain, seeming to depend on polarization, wind direction, and grazing angle in often confusing ways. Yet various empirical descriptions and statistical characterizations are available that allow much of the useful sea clutter regime to be described in ways that can be of practical value to the radar community, provided that some care is taken in defining and observing the pertinent parameters. However, there are major areas of uncertainty present at any wind speed, whenever the grazing angle goes below a few degrees and the surface illumination begins to feel the effects of refraction and diffraction, and at *any* grazing angle, whenever the wind speed is less than about 10 kt, where peculiarities and uncertainties in the generation of surface roughness begin to emerge most strongly. At the low grazing angles encountered in maritime radar operations, sea clutter becomes spiky and intermittent, requiring special attention to signal processing and the interpretation of the radar signal. Moreover, features of the sea environment such as rain, currents, slicks, and refractive anomalies can confuse the reliable separation of target returns from clutter artifacts.

The question of microwave sea clutter theory remains unsettled. The most popular model, the two-scale Bragg model, is actually an assemblage of assumptions supported by circumstantial evidence; there is still no clear reason why it should work when it does. In fact, there is increasing evidence from both the tank and the open ocean that this model fails to account for many aspects of measured sea scatter behavior. Augmenting it with a term expressing the effect of breaking waves in active seas has improved predictions, but still perpetuates the ad hoc character of composite-surface models. Theories based on scattering by surface features have begun to show promise, and at least one of these features—the breaking wave at various scales, macro to micro—is increasingly recognized as an important contributor to

sea clutter for low grazing angles and short pulses in particular. The major problem of characterizing these features in a manner useful to quantitative predictions is still being addressed. Perhaps the expression of the scattering properties of the surface in terms of an intrinsic surface property like its fine-scale curvature might introduce an organizing principle for the many bits and pieces that presently make up the theory of sea scatter.

In the last edition of this Handbook (1990), the final statement in this chapter was "It will be interesting to see what progress will have been made in the theory of sea clutter by the publication of the next edition of this handbook." It appears that the answer is—not much. But there is some evidence that the theory of sea scatter is gradually being freed from the paralyzing monotheism of Bragg scatter, so there is hope for the future.

REFERENCES

1. D. E. Kerr, *Propagation of Short Radio Waves*, New York: McGraw-Hill Book Company, 1951.
2. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill Book Company, 2001.
3. F. E. Nathanson, *Radar Design Principles*, 2nd Ed., New York: McGraw-Hill Book Company, 1991.
4. M. W. Long, *Radar Reflectivity of Land and Sea*, 3rd Ed., Norwood, MA: Artech House, 2001.
5. D. Crombie, "Doppler spectrum of sea echo at 13.56 Mc/s," *Nature*, vol. 175, pp. 681–683, 1955.
6. J. W. Wright, "A new model for sea clutter," *IEEE Trans*., vol. AP-16, pp. 217–223, 1968.
7. F. G. Bass, I. M. Fuks, A. I. Kalmykov, I. E. Ostruvsky, and A. D. Rosenberg, "Very high frequency radio wave scattering by a disturbed sea surface," *IEEE Trans*., vol. AP-16, pp. 554–568, 1968.
8. D. Barrick and Q. Peake, "A review of scattering from surfaces with different roughness scales," *Radio Sci*., vol. 3, pp. 865–868, 1968.
9. D. Atlas, R. C. Beal, R. A. Brown, P. De Mey, R. K. Moore, C. G. Rapley, and C. T. Swift, "Problems and future directions in remote sensing of the oceans and troposphere: a workshop report," *J. Geophys. Res*., vol. 9(C2), pp. 2525–2548, 1986.
10. D. Holiday, G. St-Cyr, and N. E. Woods, "A radar ocean imaging model for small to moderate incidence angles," *Int. J. Remote Sensing*, vol. 7, pp. 1809–1834, 1986.
11. D. S. Kwoh and B. M. Lake, "A deterministic, coherent, and dual-polarized laboratory study of microwave backscattering from water waves, part 1: Short gravity waves without wind," *IEEE J. Oceanic Eng*., vol. OE-9, pp. 291–308, 1984.
12. L. B. Wetzel, "Electromagnetic scattering from the sea at low grazing angles," in *Surface Waves and Fluxes: Current Theory and Remote Sensing*, chap. 12, G. L. Geernaert and W. J. Plant (eds.), Dordrecht, Netherland: Reidel, 1989.
13. L. B. Wetzel, "A time domain model for sea scatter," *Radio Sci*., vol. 28, no. 2, pp. 139–150, March–April 1993.
14. D. Middleton and H. Mellin, "Wind-generated solutions, A potentially significant mechanism in ocean surface wave generation and wave scattering," *IEEE J. Oceanic Eng*., vol. OE-10, pp. 471–476, 1985.
15. S. Tang and O. H. Shemdin, "Measurement of high-frequency waves using a wave follower," *J. Geophys. Res*., vol. 88, pp. 9832–9840, 1983.
16. W. J. Pierson and L. Moskowitz, "A proposed spectral form for fully developed seas based on the similarity theory of S. A. Kitaigorodskii," *J. Geophys. Res*., vol. 69, pp. 5181–5190, 1964.

17. O. M. Phillips, "Spectral and statistical properties of the equilibrium range in wind-generated gravity waves," *J. Fluid Mech*., vol. 156, pp. 505–531, July 1985.
18. W. J. Pierson, Jr. and M. A. Donelan, "Radar scattering and equilibrium ranges in wind-generated waves with application to scatterometry," *J. Geophys. Res*., vol. 91(C5), pp. 4971–5029, 1987.
19. S. A. Kitaigorodskii, "On the theory of the equilibrium range in the spectrum of wind-generated gravity waves," *J. Phys. Oceanogr*., vol. 13, pp. 816–827, 1983.
20. B. Kinsman, *Wind Waves*, Englewood Cliffs, NJ: Prentice-Hall, 1965.
21. O. M. Phillips, F. L. Posner, and J. P. Hansen, "High range resolution radar measurements of the speed distribution of breaking events in wind generated waves: Surface impulse and wave energy dissipation rates," *J. Phys. Oceanog*., vol. 21, 450, 2001.
22. J. Wu, "Variations of whitecap coverage with wind stress and water temperature," *J. Phys Oceanogr*., vol. 18, pp. 1448–1453, October 1988.
23. O. M. Phillips, "Radar returns from the sea surface–Bragg scattering and breaking waves," *J. Phys. Oceanogr*., vol. 18, pp. 1063–1074, 1988.
24. W. P. Plant, "A new interpretation of sea surface slope probability density functions," *J. Geophys. Res*., vol. 108, no. C9, p. 3295, 2003.
25. J. C. Daley, J. T. Ransone, J. A. Burkett, and J. R. Duncan, "Sea clutter measurements on four frequencies," Naval Res. Lab. Rept. 6806, November 1968.
26. G. R. Valenzuela and R. Laing, "On the statistics of sea clutter," Naval Res. Lab. Rept. 7349, December 1971.
27. N. W. Guinard, J. T. Ransone, Jr., and J. C. Daley, "Variation of the NRCS of the sea with increasing roughness," *J. Geophys. Res*., vol. 76, pp. 1525–1538, 1971.
28. J. C. Daley, "Wind dependence of radar sea return," *J. Geophys. Res*., vol. 78, pp. 7823–7833, 1973.
29. K. D. Ward, C. J. Baker, and S. Watts, "Maritime surveillance radar Part 1: Radar scattering from the ocean surface," *IEE Proc*., vol. 137, Pt F, no. 2, April 1990.
30. S. Watts, K. D. Ward, and R.T. A. Tough, "The physics and modeling of discrete spikes in radar sea clutter," presented at 2005 IEEE International Radar Conference, 2005.
31. H. Masuko, K. Okamoto, M. Shimada, and S. Niwa, "Measurement of microwave backscattering signatures of the ocean surface using X band and K_a band airborne scatterometers," *J. Geophys. Res*., vol. 91(C11), pp. 13065–13083, 1986.
32. A. I. Kalmykov and V. V. Pustovoytenko, "On Polarization features of radio signals scattered from the sea surface at small grazing angles," *J. Geophys. Res*., vol. 81, pp. 1960–1964, 1976.
33. I. Katz and L. M. Spetner, "Polarization and depression angle dependence of radar terrain return," *J. Res. Nat. Bur. Stand., Sec. D*. vol. 64-D, pp. 483–486, 1960.
34. F. Feindt, V. Wismann, W. Alpers, and W. C. Keller, "Airborne measurements of the ocean radar cross section at 5.3 GHz as a function of wind speed," *Radio Sci*., vol. 21, pp. 845–856, 1986.
35. L. C. Schroeder, P. R. Schaffner, J. L. Mitchell, and W. L. Jones, "AAFE RADSCAT 13.9-GHz measurements and analysis: Wind-speed signature of the ocean," *IEEE J. Oceanic Eng*., vol. OE-10, pp. 346–357, 1985.
36. G. P. de Loor and P. Hoogeboom, "Radar backscatter measurements from Platform Noordwijk in the North Sea," *IEEE J. Oceanic Eng*., vol. OE-7, pp. 15–20, January 1982.
37. A. H. Chaudhry and R. K. Moore, "Tower-based backscatter measurements of the sea," *IEEE J. Oceanic Eng*., vol. OE-9, pp. 309–316, December 1984.
38. F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing, Active and Passive*, Vol. III, Reading, MA: Addison-Wesley Publishing Company, 1986, Sec. 20.2.
39. B. Spaulding, D. Horton, and P. Huong, "Wind Aspect Factor in Sea Clutter Modeling," presented at 2005 IEEE International Radar Conference, 2005.
40. B. L. Lewis and I. D. Olin, "Experimental study and theoretical model of high-resolution backscatter from the sea," *Radio Sci*., vol. 15, pp. 815–826, 1980.
41. J. P. Hansen and V. F. Cavaleri, "High resolution radar sea scatter, experimental observations and discriminants," Naval Research Laboratory Report No. 8557, 1982.

42. D. Trizna, "Measurement and interpretation of North Atlantic Ocean Marine Radar Sea Scatter," Naval Res. Lab. Rept. 9099, May 1988.
43. P. Beckmann, *Probability in Communication Engineering*, New York: Harcourt, Brace and World, Inc., 1967, Sect. 6.2.
44. G. V. Trunk, "Radar properties of non-Rayleigh sea clutter," *IEEE Trans*., vol. AES-8, pp. 196–204, 1972.
45. P. H. Y. Lee, J. D. Barter, K. L. Beach, C. L. Hindman, B. M. Lake, H. Rungaldier, J. C. Shelton, A. B. Williams, R. Yee, and H. C. Yuen, "X band microwave scattering from ocean waves," *J. Geophys. Res*., vol. 100, no. C2, pp. 2591–2611, February 1995.
46. M. Katzin, "On the mechanisms of radar sea clutter," *Proc. IRE*, vol. 45, pp. 44–45, January 1957.
47. L. B. Wetzel, "A model for sea backscatter intermittency at extreme grazing angles," *Radio Sci*., vol. 12, pp. 749–756, 1977.
48. I. M. Hunter and T. B. A. Senior, "Experimental studies of sea surface effects on low angle radars," *Proc. IEE*, vol. 113, pp. 1731–1740, 1966.
49. H. Sittrop, "X- and K_u-band radar backscatter characteristics of sea clutter," in *Proc. URSI Commission II Specialist Meeting on Microwave Scattering from the Earth, E*. Schanda (ed.), Bern, 1974.
50. B. G. Smith, "Geometrical shadowing of the random rough surface," *IEEE Trans*., vol. AP-15, pp. 668–671, 1967.
51. F. B. Dyer and N. C. Currie, "Some comments on the characterization of radar sea echo," in *Dig. Int. IEEE Symp. Antennas Propagat*., July 10–12, 1974.
52. D. E. Barrick, J. M. Headrick, R. W. Bogle, and D. D. Crombie, "Sea backscatter at HF: Interpretation and utilization of the echo," *Proc. IEEE*, vol. 62, 1974.
53. C. C. Teague, G. L. Tyler, and R. H. Stewart, "Studies of the sea using HF radio scatter," *IEEE J. Oceanic Eng*., vol. OE-2, pp. 12–19, 1977.
54. J. C. Wiltse, S. P. Schlesinger, and C. M. Johnson, "Back-scattering characteristics of the sea in the region from 10 to 50 KMC," *Proc. IRE*, vol. 45, pp. 220–228, 1957.
55. G. W. Ewell, M. M. Horst, and M. T. Tuley, "Predicting the performance of low-angle microwave search radars—Targets, sea clutter, and the detection process," *Proc. OCEANS 79*, pp. 373–378, 1979.
56. W. K. Rivers, "Low-angle radar sea return at 3-mm wavelength," Final Tech. Rept., Georgia Institute of Technology, Engineering Experiment Station, Contract N62269-70-C-0489, November 1970.
57. L. B. Wetzel, "On microwave scattering by breaking waves," in *Wave Dynamics and Radio Probing of the Ocean Surface*, Chap. 18, O. M. Phillips and K. Hasselmann (eds.), New York: Plenum Press, 1986, pp. 273–284.
58. B. L. Hicks, N. Knable, J. Kovaly, G. S. Newell, J. P. Ruina, and C. W. Sherwin, "The spectrum of X-band radiation backscattered from the sea surface," *J. Geophys. Res*., vol. 65, pp. 825–837, 1960.
59. G. R. Valenzuela and R. Laing, "Study of doppler spectra of radar sea echo," *J. Geophys. Res*., vol. 65, pp. 551–562, 1970.
60. V. W. Pidgeon, "Doppler dependence of sea return," *J. Geophys. Res*., vol. 73, pp. 1333–1341, 1968.
61. Y. U. Mel'nichuk and A. A. Chernikov, "Spectra of radar signals from sea surface for different polarizations," *Izv. Atmos. Oceanic. Phys*., vol. 7, pp. 28–40, 1971.
62. J. W. Wright and W. C. Keller, "Doppler spectra in microwave scattering from wind waves," *Phys. Fluids*, vol. 14, pp. 466–474, 1971.
63. D. Trizna, "A model for doppler peak spectral shift for low grazing angle sea scatter," *IEEE J. Oceanic Eng*., vol. OE-10, pp. 368–375, 1985.
64. P. H. Y. Lee, J. D. Barter, K. L. Beach, E. Caponi, C. L. Hindman, B. M. Lake, H. Rungaldier, and J. C. Shelton, "Power spectral lineshapes of microwave radiation backscattered from sea surfaces at small grazing angles," *IEE Proc.-Radar, Sonar Navig*., vol. 142, no. 5, pp. 252–258, October 1995.
65. F. E. Nathanson, *loc cit*, Sec. 7.5.

66. W. C. Keller, W. J. Plant, and G. R. Valenzuela, "Observation of breaking ocean waves with coherent microwave radar," in *Wave Dynamics and Radio Probing of the Ocean Surface*, chap. 19, O. M. Phillips and K. Hasselmann (eds.), New York: Plenum Press, 1986, pp. 285–292.
67. R. K. Moore, Y. S. Yu, A. K. Fung, D. Kaneko, G. J. Dome, and R. E. Werp, "Preliminary study of rain effects on radar scattering from water surfaces," *IEEE J. Oceanic Eng*., vol. OE-4, pp. 31–32, 1979.
68. R. F. Contreras, W. J. Plant, W, C. Keller, K. Hayes, and J. Nystuen, "Effects of rain on Ku band backscatter from the ocean," *J. Geophys. Res*., vol. 108, no. C5, pp. 3165–3180, 2003.
69. J. P. Hansen, "High resolution radar backscatter from a rain disturbed sea surface," presented at ISNR-84 Rec., Tokyo, October 22–24, 1984.
70. J. P. Hansen, "A system for performing ultra high resolution backscatter measurements of splashes," in *Proc. Int. Microwave Theory & Techniques Symp*., Baltimore, 1986.
71. L. B. Wetzel, "On the theory of electromagnetic scattering from a raindrop splash," *Radio Sci*., vol. 25, No. 6, pp. 1183–1197, 1990.
72. R. Romeiser, A. Schmidt, and W. Alpers, "A three-scale composite surface model for the ocean wave-radar modulation transfer function," *J. Geophys. Res*., vol. 99, pp. 9785–9801, 1994.
73. B. LeMehaute and T. Khangaonkar, "Dynamic interaction of intense rain with water waves," *J. Phys. Oceanog*., vol. 20, December 1990.
74. L. B. Wetzel, "On the origin of long-period features in low-angle sea backscatter," *Radio Sci*., vol. 13, pp. 313–320, 1978.
75. P. Gerstoft, L. T. Rogers, W. S. Hodgkiss, and L. J. Wagner, "Refractivity estimation using multiple elevation angles," *IEEE J. of Oceanic Eng*., vol. 28, no. 3, pp. 513–525, July 2003.
76. D. E. Barrick, "Near-grazing illumination and shadowing of rough surfaces," *Radio Sci*., vol. 30, no. 3, pp. 563–580, May-June 1995.
77. J. M. Sturm and J. C. West, "Numerical study of shadowing in electromagnetic scattering from rough dielectric surfaces," *IEEE Tran. in Geosci. and Remote Sensing*, vol. 36, no. 5, September 1998.
78. R. B. Perry and G. R. Schimke, "Large-amplitude internal waves observed off the northwest coast of Sumatra," *J. Geophys. Res*., vol. 70, pp. 2319–2324, 1965.
79. W. Alpers and I. Hennings, "A theory of the imaging mechanism of underwater bottom topography by real and synthetic aperture radar," *J. Geophys. Res*., vol. 89. pp. 10529–10546, 1984.
80. W. Garrett, "Physicochemical effects of organic films at the sea surface and their role in the interpretation of remotely sensed imagery," in *ONRL Workshop Proc.—Role of Surfactant Films on the Interfacial Properties of the Sea Surface*, F. L. Herr and J. Williams (eds.), November 21, 1986, pp. 1–18.
81. H. Huhnerfuss, W. Alpers, W. D. Garrett, P. A. Lange, and S. Stolte, "Attenuation of capillary and gravity waves at sea by monomolecular organic surface films," *J. Geophys. Res*., vol. 88, pp. 9809–9816, 1983.
82. J. C. Scott, "Surface films in oceanography," in *ONRL Workshop Proc.—Role of Surfactant Films on the Interfacial Properties of the Sea Surface*, F. L. Herr and J. Williams (eds.), November 21, 1986, pp. 19–40.
83. N. W. Guinard, "Radar detection of oil spills," presented at Joint Conf. Sensing of Environmental Pollutants, AIAA Pap. 71–1072, Palo Alto, CA, November 8–10, 1971.
84. H. Hühnerfuss, W. Alpers, A. Cross, W. D. Garrett, W. C. Keller, P. A. Lange, W. J. Plant, F. Schlude, and D. L. Schuler, "The modification of X and L band radar signals by monomolecular sea slicks," *J. Geophys. Res*., vol. 88, pp. 9817–9822, 1983.
85. C. S. Cox and W. H. Munk, "Statistics of the sea surface derived from sun glitter," *J. Mar. Res*., vol. 13, pp. 198–227, 1954.
86. W. Alpers and H. Hühnerfuss, "Radar signatures of oil films floating on the sea surface and the Marangoni effect," *J. Geophys. Res*., vol. 93, pp. 3642–3648, April 15, 1988.
87. H. Masuko and H. Inomata, "Observations of artificial slicks by X and K_a band airborne scatterometers," in *Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS'88)*, Edinburgh, September 12–16, 1988, pp. 1089–1090. Published by Noordwijk, Netherlands: European Space Agency, ESTEC, 1988.

88. H. Goldstein, "Frequency dependence of the properties of sea echo," *Phys. Rev*., vol. 70, pp. 938–946, 1946.
89. A. H. Schooley, "Some limiting cases of radar sea clutter noise," *Proc. IRE*, vol. 44, pp. 1043–1047, 1956.
90. W. S. Ament, "Forward and backscattering by certain rough surfaces," *Trans. IRE*, vol. AP-4, pp. 369–373, 1956.
91. V. Twersky, "On the scattering and reflection of electromagnetic waves by rough surfaces," *Trans. IRE*, vol. AP-5, pp. 81–90, 1957.
92. D. R. Lyzenga, A. L. Maffett, and R. A. Schuchman, "The contribution of wedge scattering to the radar cross section of the ocean surface," *IEEE Trans*. vol. GE-21, pp. 502–505, 1983.
93. L. B. Wetzel, "A minimalist approach to sea backscatter—the wedge model," in *URSI Open Symp. Wave Propagat.: Remote Sensing and Communication*, University of New Hampshire, Durham, preprint volume, July 28–August 1, 1986, pp. 3.1.1–3.1.4.
94. S. O. Rice, "Reflection of electromagnetic waves from slightly rough surfaces," *Commun. Pure Appl. Math*., vol. 4, pp. 361–378, 1951.
95. W. H. Peake, "Theory of radar return from terrain," in *IRE Nat. Conv. Rec*., vol. 7, 1959, pp. 27–41.
96. G. R. Valenzuela, "Depolarization of EM waves by slightly rough surfaces," *IEEE Trans*., vol. AP-15, pp. 552–559, 1967.
97. F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough Surfaces*, New York: Pergamon Press, 1979.
98. C. Eckart, "The scattering of sound from the sea surface," *J. Acoust. Soc. Am*., vol. 25, pp. 566–570. 1953.
99. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, New York: Macmillan Company, 1963.
100. L. B. Wetzel, "HF sea scatter and ocean wave spectra," presented at URSI Spring Meet., National Academy of Sciences, Washington, April 1966.
101. A. K. Fung and G. W. Pan, "A scattering model for perfectly conducting random surfaces: I. model development," *Int. J. Remote Sensing*, vol. 8, no. 11, pp. 1579–1593, 1987.
102. G. R. Valenzuela, "Theories for the interaction of electromagnetic and oceanic waves—a review," *Boundary-Layer Meteorol*., vol. 13, pp. 61–85, 1978.
103. B. F. Kuryanov, "The Scattering of sound at a rough surface with two types of irregularity," *Sov. Phys. Acoust*., vol. 8, pp. 252–257, 1963.
104. W. J. Plant, "Bragg scattering of electromagnetic waves from the air/sea interface," in *Surface Waves and Fluxes: Current Theory and Remote Sensing*, chap. 12, G. L. Geernaert and W. J. Plant (eds.), Dordrecht, Netherlands: Reidel, 1988.
105. A. Schmidt, V. Wismann, R. Romeiser, and W. Alpers, "Simultaneous measurements of the ocean wave radar modulation transfer function at L, C and X bands from the research platform Nordsee," *J. Geophys. Res*., vol. 100, pp. 8815–8827, 1995.
106. W. P. Plant, "A stochastic, multiscale model of microwave backscatter from the ocean," *J. Geophys. Res*., vol. 107, no. C9, p. 3120, 2002.
107. V. Kudryavtsev, D. Hauser, G. Caudal, and B. Chapron, "A semiempirical model of the normalized radar cross-section of the sea surface 1. Background model," *J. Geophys. Res*., vol. 108, no. C3, 8054, 2003.
108. V. Kudryavtsev, D. Hauser, G. Caudal, and B. Chapron, "A semiempirical model of the normalized radar cross-section of the sea surface 2. Radar modulation transfer function," *J. Geophys. Res*., vol. 108, no. C3, 8055, 2003.
109. E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, 2nd Ed., Boston: Artech House, 1993.
110. M. S. Longuet-Higgins and J. S. Turner "An 'entraining plume' model of a spilling breaker," *J. Fluid Mech*., vol. 63, pp. 1–20, 1974.

111. K. J. Sangston, "Toward a theory of ultrawideband sea scatter," in *IEEE National Radar Conference 1997*, May 13–15, 1997, pp. 160–165.

112. D. Trizna (private communication), 1989.

113. J. C. West and M. A. Sletten, "Multipath EM scattering from breaking waves at grazing incidence," *Radio Sci*., vol. 32, no. 3, pp. 1455–1467, 1997.

114. M. A. Sletten and J. C. West, "Radar investigations of breaking water waves at low grazing angles with simultaneous high-speed optical imagery," *Radio Sci*., vol. 38, no. 6, p. 1110, 2003.

115. J. Chen, T. Lo, J. Litva, and H. Leung, "Scattering of electromagnetic waves from a time-varying fractal surface," *Microwave and Optical Tech. Lett*., vol. 6, no. 1, p. 87, 2003.

116. S. Haykin, "Radar clutter attractor: implications for physics, signal processing and control," *IEE Proc. Radar: Sonar Navig*., vol. 146, no. 4, p. 177, August 1999.

117. R. Harrington, *Time-Harmonic Electromagnetic Fields*, New York: McGraw-Hill, 1961.

118. D. J. Donohue, H-C Ku, D. R. Thompson, and J. Sadowski, "Direct numerical simulation of electromagnetic rough surface and sea scattering by an improved banded matrix iterative method, "*Johns Hopkins APL Tech. Digest*, vol. 18, no. 2, pp. 204–215, 1997.

119. C. L. Rino and H. D. Ngo, "Numerical simulation of low-grazing-angle ocean microwave backscatter and its relation to sea spikes," *IEEE PGAP*, vol. 46, no. 1, 133–141, 1998.

120. J. C. West, J. M. Sturm, and A-J Ja: Low-Grazing Scattering from Breaking Water Waves Using an Impedance Boundary MM/GTD Approach, *IEEE Trans. Antennas Propagat*., vol. 46, no. 1, pp. 93–100, January 1998.

121. D. B. Coakley, P. M. Haldeman, D. G. Morgan, K. R. Nicolas, D. R. Penndorf, L. B. Wetzel, and C. S. Weller, "Electromagnetic scattering from large steady breaking waves," *Experiments in Fluids*, vol. 30, no. 5, pp. 479–487, May 2001.

122. M. R. Keller, B. L. Gotwols, W. J. Plant, and W. C. Keller, "Comparison of optically-derived spectral densities and microwave cross-sections in a wind-wave tank," *J. Geophys., Res*., vol. 100, no. C8, pp. 16163–16178, 1995.

123. E. A. Ericson, D. R. Lyzenga, and D. T. Walker, "Radar backscatter from stationary breaking waves," *J. Geophys. Res*., vol. 104, Issue C12, p. 29679, 1999.

124. Y. G. Trokhimovski, "Gravity–capillary wave curvature spectrum and mean-square slope retrieved from microwave radiometric measurements (coastal ocean probing experiment)," *J. f Atmosph. Oceanic Tech*., vol. 17, no. 9, pp. 1259–1270, 2000.

Chapter 16

Ground Echo

Richard K. Moore

The University of Kansas

16.1 INTRODUCTION

Radar ground return is described by σ^0, the differential scattering cross section, or scattering coefficient (scattering cross section per unit area), rather than by the total scattering cross section σ used for discrete targets.[1] Since the total cross section σ of a patch of ground varies with the illuminated area that is determined by the geometric radar parameters (pulse width, beamwidth, etc.), σ^0 was introduced to obtain a coefficient independent of these parameters.

Use of a differential scattering cross section implies that the return from the ground is contributed by a large number of scattering elements whose phases are independent. This is primarily because of differences in distance that, although small fractions of total distance, are many wavelengths. Superposition of power is possible for the computation of average returns. If this condition is not applicable to a particular ground target, the differential-scattering cross-section concept has no meaning for that target. For example, a very fine-resolution radar might be able to resolve a part of a car; the smooth surfaces on the car would not be properly represented by σ^0. On the other hand, a coarser radar might look at many cars in a large parking lot, and a valid σ^0 for the parking lot could be determined.

If a region illuminated at one time by a radar contains n scattering elements and the above criterion is satisfied so that power may be added, the radar equation for average power becomes

$$\bar{P}_r = \sum_i^n \frac{P_{\text{ti}} G_{\text{ti}} A_{\text{ri}} \sigma_i}{(4\pi R_i^2)^2} = \sum_i^n \frac{P_{\text{ti}} G_{\text{ti}} A_{\text{ri}} (\sigma_i / \Delta A_i) \Delta A_i}{(4\pi R_i^2)^2}$$

Here, ΔA_i is an element of surface area; and P_{ti} (power transmitted toward point i), G_{ti} (gain in direction of point i), and A_{ri} (receiving equivalent aperture in direction of point i) are values of P_t, G_t, and A_r appropriate for an element at the location of ΔA_i. The factor in parentheses in the numerator of the right-hand expression is the incremental scattering cross section for element i, but this concept is meaningful only in an average. Thus, the average power returned is given by

$$\bar{P}_r = \sum_i^n \frac{P_{\text{ti}} G_{\text{ti}} A_{\text{ri}} \sigma^0 \Delta A_i}{(4\pi R_i^2)^2}$$

Here, σ^0 has been used to denote the average value of $\sigma_i/\Delta A_i$. In this formulation, we may pass in the limit from the finite sum to the integral given by

$$\bar{P}_r = \frac{1}{(4\pi)^2} \int_{\text{Illuminated area}} \frac{P_t G_t A_r \sigma^0 \, dA}{R^4} \tag{16.1}$$

This integral is not really correct, for there is a minimum size for real, independent scattering centers. Nevertheless, the concept is widely used and is applicable as long as the illuminated area is large enough to contain many such centers.

Figure 16.1 illustrates the geometry associated with Eq. 16.1. Note that for a rectangular pulse, P_t is either zero or the peak transmitter power; but for other pulse shapes, the variation with t (or R) is significant. Actual pulses are often approximated by rectangular pulses with widths equal to their half-power widths. Real pulses cannot be rectangular after passing through real transmitter, antenna, and receiver bandwidths. The transmitting-antenna gain and receiving-antenna aperture are functions of the elevation and azimuth angles:

$$G_t = G_t(\theta, \phi) \qquad A_r = A_r(\theta, \phi) \tag{16.2a}$$

The differential scattering cross section itself is a function of both *look angle* (θ, ϕ) and ground location:

$$\sigma^0 = \sigma^0 \, (\theta, \phi, \text{location}) \tag{16.2b}$$

The integral of Eq. 16.1 must be inverted when σ^0 is measured. With narrow beams and short pulses, the inversion is relatively easy, but with the wider beams and longer pulses used in many measurements, the values obtained are sometimes poorly defined.

Some authors[2] use a scattering cross section per unit projected area rather than per unit ground area. Figure 16.2 illustrates by using a *side view* the difference

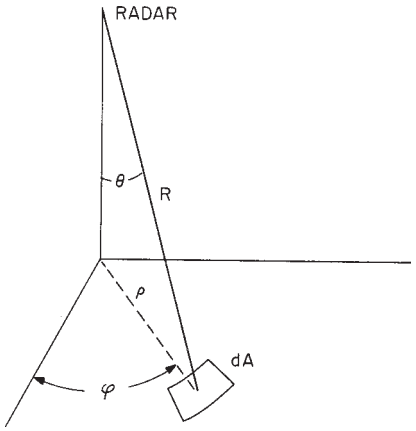


FIGURE 16.1 Geometry of the radar equation

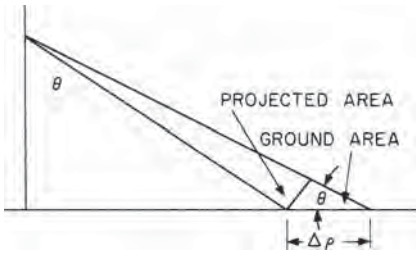


FIGURE 16.2 Ground area and projected area

between ground area and projected area. The ground area is proportional to $\Delta\rho$, and the projected area is smaller. Thus,

$$\sigma^0 A = \gamma\, d(\text{projected area}) = \gamma \cos\theta\, dA$$

or

$$\sigma^0 = \gamma \cos\theta \tag{16.3}$$

Since both γ and σ^0 are called scattering coefficients, readers of the literature must be especially careful to determine which is being used by a particular author.

Radar astronomers use a different σ[3]:

$$\frac{\text{total return power from entire surface}}{\text{power returned from perfect isotropic sphere of same radius}} \tag{16.4}$$

The resulting value for σ is usually much smaller than σ^0 for the planet at vertical incidence and is larger than the values of σ^0 near grazing incidence (return from the limb of the planet).

Relative Importance of Theory and Empiricism. The theory of radar ground return has been the subject of many publications.[4,5] The various theories, insofar as they can be confirmed by experiment, provide bases for judging the effects of variations in the dielectric properties of the ground, of the roughness of the ground and nature of vegetative or snow cover, of radar wavelength, and of angle of incidence. Viewed as aids to insight, radar ground-return theories can be extremely useful.

The validity of any ground-return theory must depend on the mathematical model used to describe the surface, as well as on the approximations required to obtain answers. Even the simplest ground surface, the sea, is extremely difficult to describe accurately; it is homogeneous to beyond the skin depth, contains relatively modest slopes, and (except for spray) has no part above another part of the surface. At grazing angles, shadowing of one wave by another might occur. Land surfaces are much more difficult to describe: Imagine an adequate mathematical description of the shape of a forest (when every leaf and pine needle must be described). Furthermore, land surfaces are seldom homogeneous either horizontally or with depth.

Since a true mathematical description of the ground surface appears out of the question, empirical measurements are necessary to describe the radar return from natural surfaces. The role of theory is to aid in interpreting these measurements and to suggest how they may be extrapolated.

Available Scattering Information. Prior to 1972, the lack of coordinated research programs over the necessary long period resulted in only one really usable set of measurements, developed at Ohio State University.[2,6] Since that time, extensive measurements have been made from trucks and helicopters by the University of Kansas,[8,9] groups in the Netherlands,[10,11] and several groups in France.[12,13,14] These measurements concentrated especially on vegetation, with the Kansas measurements also including some work on snow and extensive work on sea ice. Most of these measurements were in the 10° to 80° range of incidence angles. Measurements near vertical are scarcer.[15,16] Well-controlled experiments near grazing are also scarce except for a major Lincoln Laboratory program.[17]

Airborne measurements are necessary to make larger scattering areas accessible. Although airborne programs for special purposes have been legion, curves of scattering coefficient versus angle for a known homogeneous area are scarce. The work at the MIT Radiation Laboratory[18] was early. Work by Philco Corporation,[19] Goodyear Aerospace Corporation,[20] General Precision Laboratory,[21] and the U.S. Naval Research Laboratory (NRL)[22–25] programs were important early on. The Canada Centre for Remote Sensing (CCRS) has made numerous airborne and ground-based scatterometer measurements,[26,27] especially over sea ice. The Environmental Research Institute of Michigan (ERIM),[28] CCRS,[29] the European Space Agency (ESA),[30] and the Jet Propulsion Laboratory (JPL)[31] used imaging synthetic aperture radars (SARs) for some scattering measurements, but most were not well calibrated. Since the advent of spaceborne SARs (SIR A, B, and C; ERS 1 and 2; Radarsat; Envisat; JERS-1; and others), hundreds of papers have appeared dealing with measurements of scattering and radar applications. Moreover, the JPL AIRSAR has flown around the world, and several other airborne SARs for remote sensing have appeared in various countries, with hundreds of resulting papers. The reader should search the literature for these results, which are far too numerous to reference here. Many of the more recent SARs[32–36] also provide information on polarimetric responses.

Results of most of these measurements are summarized in Ulaby, Moore, and Fung[37] and Ulaby and Dobson.[38] More complete summaries of the earlier work and near-grazing studies are in Long[22,39] and in the work of Billingsley.[17] Many application summaries are also in the *Manual of Remote Sensing*.[40,41] Readers requiring more detailed information should consult these books.

16.2 PARAMETERS AFFECTING GROUND RETURN

Radar return depends upon a combination of system parameters and ground parameters:

Radar system parameters (Eqs. 16.1 and 16.2*a* and *b*):
- Wavelength
- Power
- Illuminated area
- Direction of illumination (both azimuth and elevation)
- Polarization (including the full polarization matrix when available)

Ground parameters:
- Complex permittivity (conductivity and permittivity)
- Roughness of surface
- Inhomogeneity of subsurface or cover to depth where attenuation reduces waves to negligible amplitude

Different wavelengths are sensitive to different elements on the surface. One of the earliest known and most striking directional effects is the *cardinal-point* effect in returns from cities: Radars looking in directions aligned with primary street grids observe stronger regular returns than radars at other angles. Horizontally polarized waves are reflected better by horizontal wires, rails, etc., than are vertically polarized waves. Vertically polarized waves are reflected better from vertical structures such as tree trunks, at least when the wavelength is comparable or larger than the trunk diameter.

If the geometry of two radar targets were the same, the returns would be stronger from the target with higher complex permittivity because larger currents (displacement or conduction) would be induced in it. Because identical geometries with differing permittivities do not occur in nature, this distinction is not easy to measure. Effective permittivity for ground targets is very strongly influenced by moisture content since the relative permittivity of liquid water is from about 60 at X band to about 80 at S band and longer wavelengths, whereas most dry solids have permittivities less than 8. Attenuation is also strongly influenced by moisture since wet materials usually have higher conductivity than the same materials dry. Figures 16.3 and 16.4 show the effect of moisture content on properties of plants and of soil. The high permittivity of plants with much moisture means that radar return from crops varies as the plants mature, even when growth is neglected.

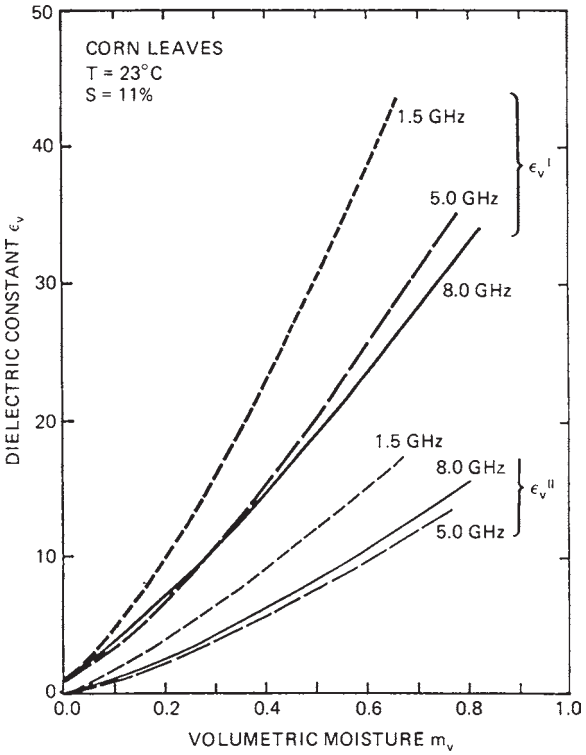


FIGURE 16.3 Measured moisture dependence of the dielectric constant of corn leaves at 1.5, 5.0, and 8.0 GHz. *S* is the salinity of water content in parts per thousand, $\varepsilon_v = \varepsilon_v' - j\varepsilon_v''$ is the complex dielectric constant in Fm^{-1}, and m_v is the volumetric moisture content in kgm^{-3}. (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

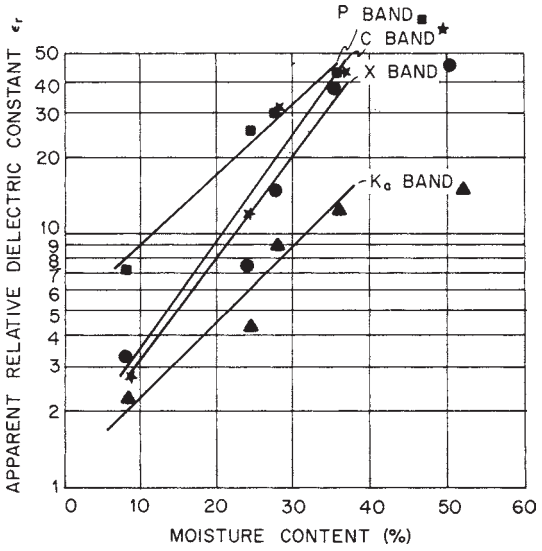


FIGURE 16.4 Apparent relative dielectric constant versus moisture content (Richfield slit loam) (*after J. R. Lundien*[42])

The roughness of surfaces (especially natural ones) is difficult to describe mathematically but easy to understand qualitatively. Thus it is easy to see that a freshly plowed field is rougher than the same field after rain and wind have been at work on it. A forest is inherently rougher than either a field or a city. It is harder to see the difference between the roughness of natural areas and the roughness of a city that has flat walls interspersed with window sills and with curbs, cars, and sidewalks.

Surfaces that are relatively smooth tend to reflect radio waves in accordance with the Fresnel-reflection direction,* so they give strong backscatter only when the look angle is nearly normal to the surfaces. Rough surfaces, on the other hand, tend to reradiate nearly uniformly in all directions, and so they give relatively strong radar returns in any direction.

The problem of radar scatter is complicated because waves penetrate significant distances into many surfaces and vegetation canopies, and internal reflection and scatter contribute to the return. Measurements of attenuation for field crops[43,44] and grasses[45] show that most of the return is from the upper layers, with some contribution by the soil and lower layers if the vegetation is not very dense. At C band and higher frequencies, most of the signal returned from trees is usually from the upper and middle branches when the trees are in leaf,[46–50] although in winter the surface is a major contributor to the signal. At L band, and especially at VHF, the signal penetrates farther, so trunks and the ground can be major contributors even when the trees are leafed out.[51]

Additional problems occur near grazing incidence.[17,52] Because of the low angle with the surface, shadowing frequently occurs—some parts of the target are obscured by intervening projections such as hills and buildings. Parts of the area that are somewhat elevated have the signal modified by the effect of multipath interference between

* Angle of reflection equals angle of incidence.

the direct ray and one reflected off the ground. Since the scattering from relatively level surfaces is very small, any projection may give a return much stronger than the background, thereby skewing the statistics so a Rayleigh distribution no longer applies to the average signal. Objects such as trees, buildings, fence posts, and power lines give localized echoes strong relative to their surroundings.

Moreover, the signal from surfaces without projections falls off very rapidly for depression angles within a few degrees of grazing. This means that the effect of small local slopes can be very significant in modulating the return signal, not just in shadowing.

16.3 THEORETICAL MODELS AND THEIR LIMITATIONS

Descriptions of a Surface. Many theoretical models for radar return from the ground assume a rough boundary surface between air and an infinite homogeneous half space. Some include either vertical or horizontal homogeneities in the ground properties and in vegetative or snow covers.

Surface descriptions suitable for use in mathematical models are necessarily greatly idealized. Few natural grounds are truly homogeneous in composition over very wide areas. Descriptions of their detailed shape must be simplified if they are to be handled analytically, although computers permit the use of true descriptions. Very few surfaces have ever been measured to the precision appropriate for centimeter-wavelength radars; even for these, there is no assurance that scattering boundaries do not exist within a skin depth beneath the surface. Surfaces containing vegetation and conglomerate rocks almost completely defy description.

Statistical descriptions of surfaces are used for most theories, since a theory should be representative of some kind of surface class, rather than of a particular surface, and since exact description is so difficult. The statistical descriptions themselves must be oversimplified, however. Many theories assume isotropic statistics, certainly not appropriate for plowed fields or gridded cities. Most theories assume some kind of model involving only two or three parameters (standard deviation, mean slope, correlation distance, etc.), whereas natural (or human-made) surfaces seldom are so simply described. The theories for vegetation and other volume scatters have more parameters. For near-grazing conditions, the models must account for shadowing.

Simplified Models. Early radar theories for ground return assumed, as in optics, that many targets could be described by a Lambert-law variation of intensity; that is, the differential scattering coefficient varies as $\cos^2\theta$, with θ being the angle of incidence. This "perfectly rough" assumption was soon found wanting, although it is a fair approximation for the return from many vegetated surfaces over the midrange of angles of incidence.

Clapp[18] described three models involving assemblies of spheres, with different spacings and either with or without a reflecting ground plane. These models yield variations from σ^0 independent of angle through $\sigma^0 \propto \cos\theta$ to $\sigma^0 \propto \cos^2\theta$. Since the sphere models are highly artificial, only the resulting scatter laws need be considered. Most targets give returns that vary more rapidly over part of the incidence-angle regime than these models, although forests and similar rough targets of some depth sometimes give such slowly varying returns.

Since these rough-surface models usually fail to explain the rise in return near vertical incidence, other simplified models combine Lambert's law and other rough-surface scattering models with specular reflection at vertical incidence, and a smooth curve is drawn between the specular value and the rough-surface prediction.

Specular reflection is defined as reflection from a smooth plane and obeys the Fresnel reflection laws.[53] At normal incidence, the specular-reflection coefficient is therefore

$$\Gamma_R = \frac{\eta_g - \eta_0}{\eta_g + \eta_0}$$

where η_0 and η_g are the intrinsic impedances of air and earth, respectively. The fraction of total incident power specularly reflected from a rough surface is[7]

$$e^{-2(2\pi\sigma_h/\lambda)^2}$$

where σ_h = standard deviation of surface height variations, and
λ = wavelength

Since this proportion is down to 13.5% when $\sigma_h = \lambda/2\pi$ and to 1.8% when $\sigma_h = \lambda/(2\pi\sqrt{2})$, significant specular reflection is seldom found for the centimeter wavelengths generally used for radar. Nevertheless, a simplified model like this is convenient for some purposes.

Observation of reflected sunlight from rippled water, from roads, and from other smooth surfaces leads to the postulation of a facet theory.[54,55] The only sunlight reaching the observer from smooth surfaces such as water is that from facets for which angle of incidence equals angle of reflection. Thus, the observed light may be described by methods of *geometric optics*.

When geometric optics is used to describe radar scatter, the surface of the ground is represented by small flat-plane segments. Radar return is assumed to occur only for facets oriented normal to the radar (normal orientation is required for backscatter so that the reflected wave returns to the source). Thus, if the slope distribution of such facets is known, the fraction normal to a given diverging beam can be established, and from this the return can be obtained. Geometric optics assumes zero wavelength, and so the results of such a theory are wavelength-independent, clearly not in accord with observation.

The facet model for radar return is extremely useful for qualitative discussions, and so modification to make it fit better with observation is appropriate. Two kinds of modification may be used, separately or jointly: considering the actual reradiation pattern of finite-size facets at finite wavelengths[55] and considering the effect of wavelength on establishing the effective number of facets.[56] Thus, the scatter from a facet may actually occur in directions other than that requiring that angle of incidence equal angle of reflection. Figure 16.5 illustrates this. For large facets (compared with wavelength), most of the return occurs almost at normal incidence, whereas for small facets the orientation may be off normal by a considerable amount without great reduction in scatter. As the wavelength is increased, the category of a given facet changes from *large* to *small;* eventually the facet is smaller than a wavelength, and its reradiation pattern shape remains almost isotropic from that point. Many facets that would be separate at, say, a 1-cm wavelength are combined at a 1-m wavelength; the result may be a transition from rough- to smooth-surface behavior. Figure 16.6*a* shows a number of facets of different sizes contributing to a radar return.

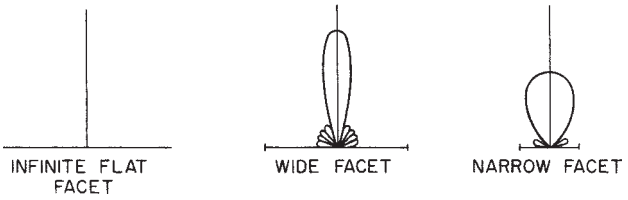


FIGURE 16.5 Normal-incidence reradiation patterns of facets

Physical Optics Models. Theories based on applications of the Kirchhoff-Huygens principle have been thoroughly developed.[37,55,57–59] The Kirchhoff approximation is that the current flowing at each point in a locally curved (or rough) surface is the same as would flow in the same surface if it were flat and oriented tangent to the actual surface. This assumption permits construction of scattered fields by assuming that the current over a rough plane surface has the same magnitude as if the surface were smooth, but with phase perturbations set by the differing distances of individual

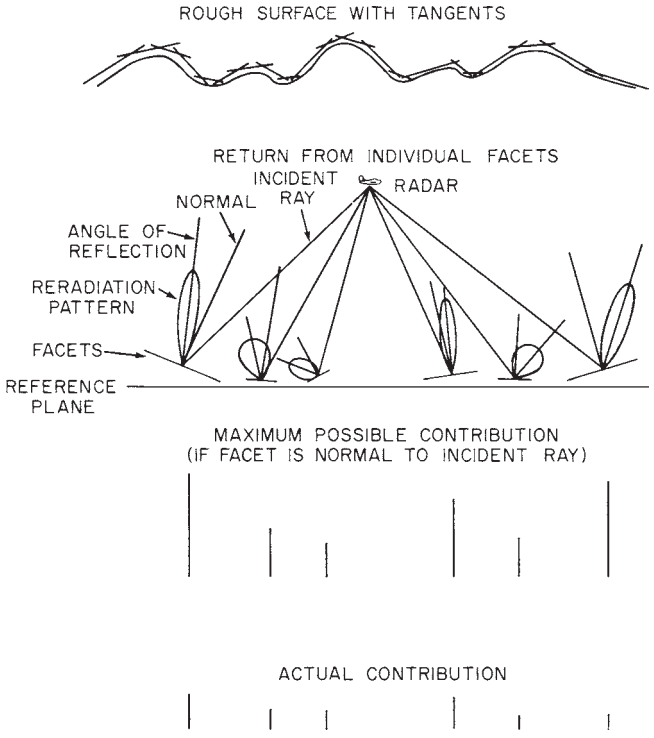


FIGURE 16.6*a* Facet model of a radar return

points from the mean plane. For surfaces assumed to be azimuthally isotropic, the usual approach yields integrals of the form

$$\frac{1}{\cos^3\theta}\int e^{-(2k\sigma_h\cos\theta)^2[1-\rho(\xi)]}J_0(2k\xi\cos\theta)\xi\,d\xi$$

where $\rho(\xi)$ = spatial autocorrelation function of surface heights
θ = angle with vertical
σ_h = standard deviation of surface heights
k = $2\pi/\lambda$
J_0 = first-order, first-kind Bessel function

The autocorrelation function of height with distance is seldom known for terrain, although it can be determined on a large scale by analysis of contour maps,[60] and it has been found for some areas by careful contouring at close intervals and subsequent analysis. Because of lack of knowledge of actual autocorrelations, most theory has been developed with artificial functions that are chosen more for their integrability than for their fit with nature; selection among them has been on the basis of which ones yield the best fit between theoretical and experimental scatter curves.

The correlation function first used[61] was gaussian:

$$\rho(\xi) = e^{-\xi^2/L^2} \tag{16.5}$$

where L is the *correlation length*. Not only is this a function that makes the integral analytically tractable, but it also gives exactly the same results as geometric optics.[62] Since it fails, like geometric optics, to explain frequency variation, it cannot be a truly representative correlation function, although it gives a scattering curve that fits several experimental curves near the vertical. The next most frequently used function is the exponential:

$$\rho(\xi) = e^{-|\xi|/L} \tag{16.6}$$

This has some basis in contour-map analysis;[60] the results fit both Earth and lunar radar return over a wider range of angles than the gaussian [60,63] (but sometimes not as well near vertical). Furthermore, it has the merit that it exhibits frequency dependence. Resulting expressions for power (scattering coefficient) variations appear in Table 16.1.

TABLE 16.1 Scattering Coefficient Variation

Correlation coefficient	Power Expression	Reference
$e^{-\xi^2/L^2}$	$\frac{K}{\sin\theta}e^{-(L^2/2\sigma_h^2)\tan^2\theta}$	Davies[61]
$e^{-\lvert\xi\rvert/L}$	$\frac{K\theta}{\cos^2\theta\sin\theta}\left(1+A\frac{\sin^2\theta}{\cos^4\theta}\right)^{-3/2}$	Voronovich[5]

Small-Perturbation and Two-Scale Models. Recognition that existing models were inadequate for describing ocean scatter led to recognition that resonance of the signal with small structures on the surface has a powerful influence on the strength of the signal received.[64,65] Thus a small-perturbation method originally proposed by Rice[66] became the most popular way to describe ocean scatter. Its application to land scatter was not far behind.

The term *Bragg scatter* is often used to describe the mechanism for the small-perturbation model. The idea comes from the concept illustrated in Figure 16.6*b*.

A single sinusoidal component of a complex surface is shown with an incoming radar wave at angle of incidence θ. The radar wavelength is λ, and the surface-component wavelength is Λ. When the signal travels an extra distance $\lambda = 2\Delta R$ between the source and two successive wave crests, the phase difference between the echoes from successive crests is 360°; so the echo signals all add in phase. If this condition is satisfied for a particular Λ and θ, it fails to be satisfied for others. Thus, this is a resonant selection for a given θ of a particular component of the surface Λ. The strength of the received signal is proportional to the height of this component and to the number of crests illuminated by the radar. If the surface has an underlying curvature, the number of illuminated crests satisfying the resonance criterion may be limited by the length of the essentially flat region; otherwise, it is limited by the radar resolution.

The theoretical expression for the scattering coefficient is[67]

$$\sigma^0_{pq} = 8k^4\sigma_1^2 \cos^4\theta \,|\, \alpha_{pq} \,|^2 \, W(2k\sin\theta, 0) \tag{16.7}$$

where p, q = polarization indices (H or V)
$k = 2\pi/\lambda$ (the radar wavenumber)
$\alpha_{HH} = R_1$ (Fresnel reflection coefficient for horizontal polarization)

$$\alpha_{VV} = (\varepsilon_r - 1)\frac{\sin^2\theta - \varepsilon_r(1+\sin^2\theta)}{[\varepsilon_r\cos\theta + (\varepsilon_r - \sin^2\theta)^{1/2}]^2}$$

where ε_r is the relative permittivity $\varepsilon' - j\varepsilon''$ and $\alpha_{VH} = \alpha_{HV} = 0$. $W(2k \sin\theta, 0)$ is the *normalized roughness spectrum* (the Fourier transform of the surface autocorrelation function). It may be written as $W(K, 0)$, where K is the wavenumber for the surface. In terms of the wavelength on the surface Λ,

$$K = 2\pi/\Lambda$$

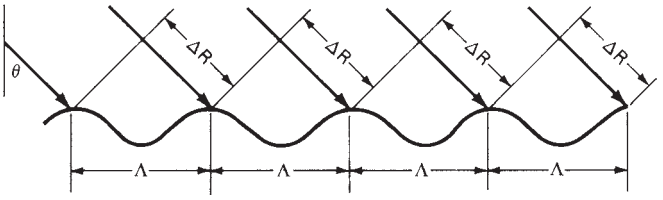


FIGURE 16.6*b* In-phase addition for Bragg scattering: $\Delta R = n\lambda/2$

Thus, the component of the surface that satisfies the Bragg resonance condition is

$$\Lambda = \lambda/2 \sin \theta \tag{16.8}$$

This means that the most important contributor to a surface return is the component of surface roughness with wavelength Λ. Even though other components may be *much* larger, the Bragg resonance makes this component more important. On the ocean, this means that tiny ripples are more important than waves that are meters high; the same applies for land-surface scatter.

As originally developed, this theory was for perturbations to horizontal flat surfaces, but it was soon modified to handle surfaces with large-scale roughness. The large-scale roughness was assumed to cause a *tilting* of the flat surface to which the small-perturbation theory could be applied. The principal problem with this approach is deciding where in the surface spectrum lies the boundary between the larger components that do the tilting and smaller components that are Bragg-resonant. Many papers have been written to describe the evolution of this theory; for a complete summary, the reader is referred to Fung's development.[68]

Other Models. The theory for volume scatter has led to many papers and continues to evolve. For a review of some of the approaches, the reader should consult Fung's summary[69] and papers by Kong, Lang, Fung, and Tsang. These models have been used reasonably successfully to describe scatter from vegetation,[70] snow,[71] and sea ice.[72] Models of straight vegetation such as wheat in terms of cylinders have had some success.[2,73] Corner-reflector effects have been used to describe strong returns from buildings at non-normal incidence angles.[74,75] Other specialized models have been used for particular purposes.

Later theoretical work for surfaces involves solving integral equations for the scattered fields.[76] This has been used both to validate other models and to better describe the true scattering from a known rough surface. This method tends to be computationally extensive. Numerical scattering computations are also in vogue.[77]

Regardless of the model used and the approach applied to determining the field strength, theoretical work only guides understanding. Actual Earth surfaces are too complex to be described adequately in any of the models, and the effects of signals that penetrate the ground and are scattered therein are too little known to permit its evaluation.

16.4 FADING OF GROUND ECHOES

The amplitude of ground echoes received by radars on moving vehicles fluctuates widely because of variations in phase shift for return from different parts of the illuminated area. In fact, even fixed radars frequently observe fluctuations in ground echoes because of motions of vegetation, wires blowing in the wind, etc. This fluctuation is referred to as *fading*.

Fading is significant for the radar engineer because one must account for the fact that a single sample of the radar return may vary widely from the mean described by σ^0. Thus, the system must be able to handle the dynamic range of fading, which may exceed 20 dB.

Regardless of the model used to describe a ground surface, signals are, in fact, returned from different positions not on a plane. As a radar moves past a patch of ground

while illuminating it, the look angle changes, and this changes the relative distances to different parts of the surface; the result is that relative phase shift is changed. This is the same kind of relative-phase-shift change with direction that is present for an antenna array and results in the antenna pattern. For ground echo, the distance is doubled, so the pattern of an echoing patch of length L has lobes of width $\lambda/2L$. This compares with λ/L for an antenna of the same cross-range length. Because the excitation of the elements of the scattering array is random, the scattering pattern in space is also random.

This fading phenomenon is usually described in terms of the doppler shift of the signal. Since different parts of the target are at slightly different angles, the signals from them experience slightly different doppler shifts. The doppler shift, of course, is simply the rate of change of phase due to motion. Thus, the total rate of change of phase for a given target is

$$\omega = \omega_c + \omega_{\text{di}} = \frac{d\phi_i}{dt} = \frac{d}{dt}(\omega_c t - 2kR_i) \tag{16.9}$$

where ω_c = carrier angular frequency
ω_{di} = doppler angular frequency for ith target
ϕ_i = phase for ith target
R_i = range from radar to ith target

The doppler shift can be expressed in terms of the velocity vector **v** as

$$\omega_{\text{di}} = -2k\frac{dR_i}{dt} = -2k\mathbf{v}\cdot\frac{\mathbf{R}_i}{R_i} = -2kv\cos(\mathbf{v},\mathbf{R}_i) \tag{16.10}$$

where bold-faced letters are vectors. Hence, the total field is given by

$$E = \sum_i A_i \exp\left\{j\left[\omega_c t - \int_0^t 2k\mathbf{v}\cdot\frac{\mathbf{R}_i}{R_i}\,dt - 2kR_{i0}\right]\right\} \tag{16.11}$$

where A_i is the field amplitude of the ith scatterer and R_{i0} is the range at time zero.

The only reason the scalar product varies for different scatterers is the different angle between the velocity vector and the direction to the scatterer. This results in a separate doppler frequency for each scatterer. If we assume the locations to be random, as most theories do, the received signal is the same as one coming from a set of oscillators with random phases and unrelated frequencies. This same model of a group of randomly phased, different-frequency oscillators is used to describe noise; *thus, the statistics of the fading signal and the statistics of random noise are the same*.

This means that the envelope of the received signal is a random variable with its amplitude described by a Rayleigh distribution. Such distributions have been measured for many ground-target echoes.[23] Although the actual distributions vary widely, no better description can be given for relatively homogeneous targets. With a Rayleigh distribution, the 90% range of fading is about 18 dB, so an individual pulse return may be anywhere in this range.

When a target is dominated by one large echo (such as a metal roof oriented to give a strong return), the distribution is better described by that for a sine wave in noise. If the large echo is considerably stronger than the mean of the remaining contributors to the return, this approaches a normal distribution about the value for the large echo. This situation is particularly common for near-grazing conditions.[17]

For reference, the two distributions are given[78]:

$$p(v)\,dv = \frac{v}{\psi_0} e^{-v^2/2\psi_0}\,dv \qquad \text{(Rayleigh)}$$

$$p(v)\,dv = \frac{v}{\psi_0^{1/2}} e^{-(v^2+a^2)/2\psi_0} I_0\left(\frac{av}{\psi_0}\right) \qquad \text{(sine wave + Rayleigh)}$$

where v = envelope voltage
ψ_0 = mean square voltage
A = sine-wave peak voltage
$I_0(x)$ = Bessel function, first kind, zero order, imaginary argument

In practice, the distribution from large targets may be more complicated than either of the simple models described. Indeed, particularly near grazing incidence, the signal is often described by a *K*, a Weibull, or a log-normal distribution.[79–81] These distributions are more often used to describe the variations between different returns from an area, rather than fading. They may be thought of as describing what happens when the area contains different σ^0s, and the distribution for each is Rayleigh. Because of this, the range of variability may be even greater than the 18 dB for a Rayleigh distribution.

Fading-Rate Computations. Doppler frequency calculation is the easiest way to find fading rates. To compute the signal amplitude returned with a particular range of doppler shifts, all signals having such shifts must be summed. This requires knowing the contours of constant doppler shift (isodops) on the scattering surface. These contours must be established for each particular geometric arrangement. A simple example is presented here: horizontal motion over a plane earth. This is typical of an aircraft in ordinary cruising flight.

Consider travel in the y direction, with z vertical and the altitude (fixed) $z = h$. Then

$$\mathbf{v} = \mathbf{1}_y v$$

$$\mathbf{R} = \mathbf{1}_x x + \mathbf{1}_y y - \mathbf{1}_z h$$

where ($\mathbf{1}_x$, $\mathbf{1}_y$, $\mathbf{1}_z$) are unit vectors. Hence,

$$v_r = \mathbf{v} \cdot \frac{\mathbf{R}}{R} = \frac{vy}{\sqrt{x^2 + y^2 + h^2}}$$

where v_r is the relative speed. Curves of constant relative speed are also curves of constant doppler shift. The equation of such a curve is

$$x^2 - y^2 \frac{v^2 - v_r^2}{v_r^2} + h^2 = 0$$

This is a hyperbola. The limiting curve for zero relative speed is a straight line perpendicular to the velocity vector. Figure 16.7 shows such a set of constant-doppler-shift contours.

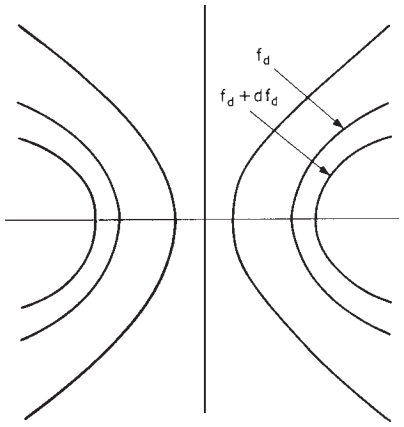


FIGURE 16.7 Contours of constant doppler frequency shift on a plane earth due to horizontal motion

The spectrum of fading can be calculated by a slight rearrangement of the radar equation (Eq. 16.1). Thus, if $W_r(f_d)$ is the power received between frequencies f_d and $f_d + df_d$, the radar equation becomes

$$W_r(f_d)\,df_d = \frac{1}{(4\pi)^2}\int_{\substack{\text{Illuminated area}\\ \text{between } f_d \text{ and } f_d+df_d}} \frac{P_t G_t A_r \sigma^0\,dA}{R^4} = \frac{df_d}{(4\pi)^2}\int \frac{P_t G_t A_r \sigma^0}{R^4}\left(-\frac{dA}{df_d}\right) \tag{16.12}$$

This is an integral in which the area element between f_d and $f_d + df_d$ is expressed in terms of coordinates along and normal to the isodops. Such coordinates must be established for each particular case.

Figure 16.8 shows the geometry for horizontal travel. The coordinate ξ is along the isodop, and η is normal to it. We can express Eq. 16.12 in terms of these coordinates as

$$W_r f_d = \frac{d\eta}{df_d}\left[\frac{\lambda^2}{(4\pi)^3}\right]\int_{\text{strip}}\left[\frac{P_t G^2 \sigma^0\,d\xi}{R^4}\right] \tag{16.13}$$

Note that P_t, the transmitted power, is nonzero in the integral only for the time it illuminates the ground. In pulse radars, only that part of the ground area providing signals back to the radar at a particular time can be considered to have finite P_t, and so the range of frequencies that can be present is limited by the pulse, as well as by the antennas and the maximum velocity.

Another example is shown in Figure 16.9. This is the small illuminated area for a narrow-beam short-pulse system with the antenna pointed straight ahead. Here, we can make linear approximations without too much error.

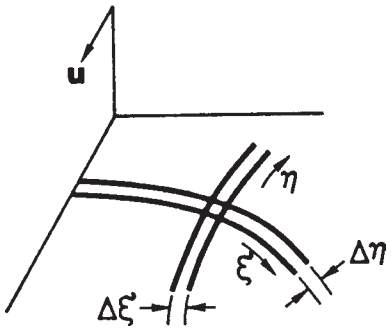


FIGURE 16.8 Geometry of complex fading calculations (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

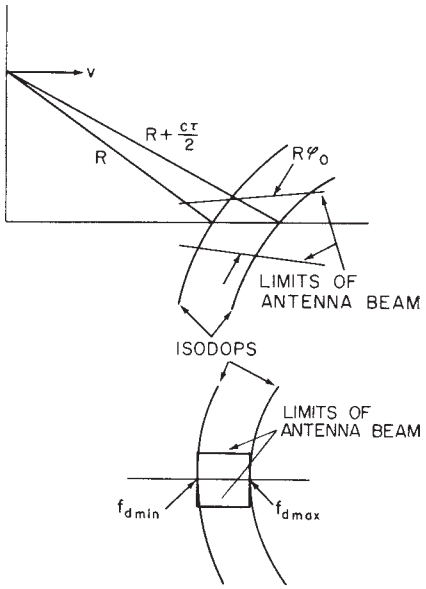


FIGURE 16.9 Geometry of doppler-shift calculations for an airborne search radar

A pulse of length τ is transmitted from an antenna of beamwidth ϕ_0. We may simplify the problem by assuming a rectangular illuminated area, $R\phi_0 c\tau/(2 \sin \theta)$. Furthermore, the curvature of the isodops may be neglected, so the doppler frequency is assumed to be the same for all maximum-range points and the same for all minimum-range points. With this assumption,

$$f_{d\max} = \frac{2v}{\lambda}\sin\theta_{\max}$$

$$f_{d\min} = \frac{2v}{\lambda}\sin\theta_{\min}$$

Thus, the total width of the doppler spectrum is

$$\Delta f_d = \frac{2v}{\lambda}(\sin\theta_{\max} - \sin\theta_{\min})$$

For short pulses and angles away from vertical, this is

$$\Delta f_d \approx \frac{2v}{\lambda}\Delta\theta\cos\theta$$

In terms of pulse length, it becomes

$$\Delta f_d = \frac{vc\tau}{2h\lambda}\frac{\cos^3\theta}{\sin\theta} \tag{16.14}$$

If the angular difference across the illuminated rectangle is small enough so that σ^0 is essentially constant, the doppler spectrum is a rectangle from $f_{\min}$ to $f_{\max}$.

In practice, antenna beams are not rectangular. The result is that the doppler spectrum for a side-looking radar like that shown in the example is not rectangular but rather has the shape of the antenna along-track pattern. Thus, if the antenna pattern in the along-track direction is $G = G(\beta)$, with β the angle off the beam center, we can express β in terms of the doppler frequency f_d as

$$\beta = f_d \lambda / 2v$$

and the spectrum is

$$W(f_d) = \frac{\lambda^3 P_t \sigma^0 r_x}{2(4\pi)^3 R^3} G^2 \left[\frac{\lambda f_d}{2v} \right]$$

where r_x is the horizontal resolution in the range direction. Of course, the half-power beamwidth may be used as an approximation, resulting in the bandwidth given by Eq. 16.13.

Effect of Detection. The effect of detecting narrowband noise has been treated extensively in the literature. Here, it is necessary only to show the postdetection spectrum of the preceding example and to consider the number of independently fading samples per second. Figure 16.10 shows the spectrum before and after square-law detection. For square-law detection, the post-detection spectrum is the self-convolution of the predetection spectrum. Only the part that passes the low-pass filters in a detector is shown in the figure. The rectangular *RF spectrum* has become a triangular *video spectrum*.

This spectrum describes the fading of the detector output for a CW radar. For a pulse radar, the spectrum is sampled by the pulse repetition frequency (PRF). If the PRF is high enough so that the entire spectrum can be reproduced (the PRF is higher than the Nyquist frequency, $2\Delta f_d$), the diagram indicated is that of the spectrum of the samples of a received pulse at a given range. Figure 16.11 shows a series of actual pulses from a moving radar, followed by a series of samples at range R_1. The spectrum shown in Figure 16.10 is the spectrum of the envelope of samples at R_1 (after low-pass filtering). The spectrum of fading at another range (or vertical angle) is different, in accord with Eq. 16.13.

For many purposes, the number of *independent* samples is important because these may be treated by using the elementary statistics of uncorrelated samples. For continuous integration, the effective number of independent sample is[78]

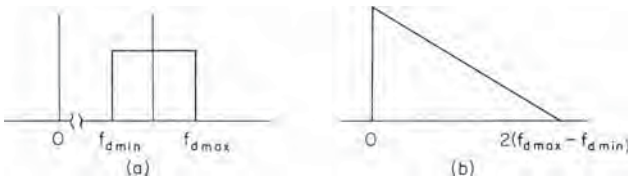


FIGURE 16.10 Spectrum of fading from a homogeneous small patch (*a*) before and (*b*) after detection

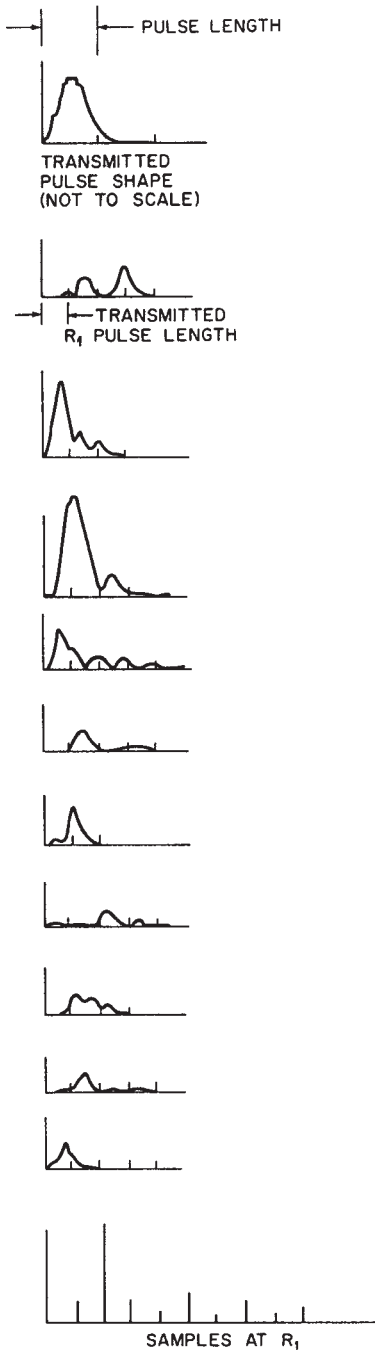


FIGURE 16.11 Fading for successive pulses of a moving radar with ground target

$$N = \frac{\overline{P_e^2} T}{2\int_0^T \left[1 - \frac{x}{T}\right] R_{sf}(x)\, dx} \tag{16.15}$$

where $\overline{P}_e$ is the mean envelope power, T is the integration (averaging) time, and $R_{sf}(t)$ is the autocovariance function for the detected voltage. For many practical purposes, if N is large, it may be approximated by

$$N \approx BT \tag{16.16}$$

where B is the effective IF bandwidth. For the effect of short integration time, see Ulaby et al.[82]

Fading samples can, of course, also be independent because motion of the vehicle causes the beam to illuminate a different patch of ground. Thus, in a particular case, the independent-sample rate may be determined either by the motion of the illuminated patch over the ground or by the doppler effect or by some combination of the two.

The number of independent samples determines the way in which the Rayleigh or other distributions may be applied. Thus, if 100 pulses give only 10 independent samples, the variance of the mean obtained by integrating these pulses is much greater than would be true if all 100 pulses were independent.

Doppler-based systems, such as doppler navigators, moving-target indicators, and synthetic-aperture radar systems, depend on the predetection spectrum for their operation, because they are coherent and do not use amplitude or square-law detection.

Moving-Target Surfaces. Sometimes clutter has internal motion. This can occur when fixed radars are used to observe movement of the sea and the land. On land, clutter motion is usually due to moving vegetation, although moving animals and machines create similar effects. The radar return from an assembly of scatterers like those shown in Figure 16.8 can change because of motion of the individual scatterers just as it changes because of motion of the radar. Thus, if each scatterer is a tree, the waving of the trees as the wind blows causes relative phase shifts between the separate scatterers; the result is fading. For a fixed radar, this may be the only fading observed, except for very slow fading due to changes in refraction. If the surface elements are stiff, they may not move enough to get significant doppler spreading, and the fading distribution may not be close to Rayleigh. See Billingsley[17] and other papers by him for more discussion of the situation for fixed radars observing ground targets. For a moving radar, this motion of the target changes the relative velocities between target element and radar so that the spectrum is different from that for a fixed surface. The width of the spectrum due to vehicle motion determines the ability of the radar to detect this target motion.

16.5 MEASUREMENT TECHNIQUES FOR GROUND RETURN

Special-purpose instrumentation radars and modified standard radars may be used to determine the ground return. Since the ground return is almost invariably due to scattering, these systems are termed *scatterometers*. Such systems may use CW signals with or without doppler processing, but they may also use both pulse and FM techniques.

Scatterometers capable of measuring response over a wide range of frequencies are called *spectrometers*.[83] Various antenna patterns from pencil beams to fan beams may be used. Systems to measure the full polarization matrix must use very careful antenna designs so that the phases of the different transmitted and received polarizations are well controlled, and leakage between polarizations is thoroughly suppressed.

CW and FM-CW Systems. The simplest scatterometer uses a stationary CW radar. Such systems are not very flexible, but they are discussed here in some detail to illustrate calibration techniques that also apply to the more complex systems.

The CW scatterometer is shown in block form in Figure 16.12. To evaluate σ^0 the ratio of transmitted to received power is required. The system depicted in Figure 16.12*a* measures transmitter power and receiver sensitivity separately. The transmitter feeds an antenna through a directional coupler so that a portion of the energy may be fed to a power meter. The receiver operates from a separate antenna (electrically isolated). The output of the receiver is detected, averaged, and digitally recorded. Its sensitivity must be checked by use of a calibration source. The calibrated signal may be fed through the receiver at a time when the transmitter is off. Figure 16.12*b* shows a similar arrangement in which the signal from the transmitter is attenuated a known amount and used to check the receiver gain. By comparing the output from the attenuated transmitter signal with that received from the ground, the scattering cross section may be determined without actually knowing the transmitted power and the receiver gain.

The calibrations shown in Figure 16.12 are incomplete without knowing the antenna patterns and absolute gains. Since accurate gain measurements are difficult, absolute calibrations may be made by comparing received signals (with proper relative calibration) from the target being measured and from a *standard target*. Standard targets may be metal spheres, Luneburg-lens reflectors, metal plates, corner reflectors, or active radar calibrators (ARCs—actually repeaters).[84,85] Of the passive calibrators, the Luneburg-lens reflector is best because it has a large cross section for its volume and has a very wide pattern so that alignment is not critical. Luneburg-lens reflectors are used for making strong radar targets of small vessels, and they may be obtained

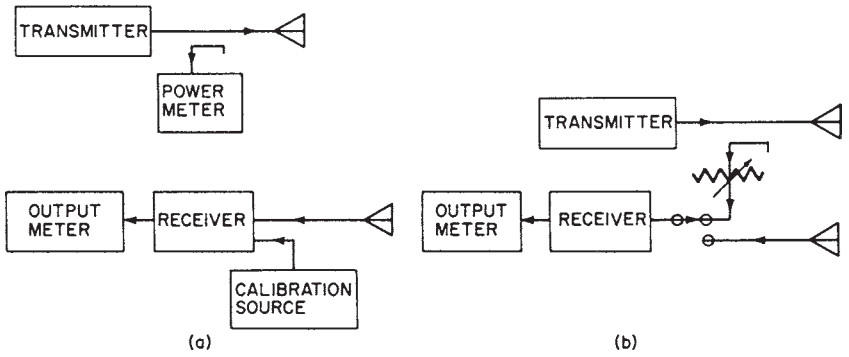


FIGURE 16.12 CW-scatterometer-system block diagram: (*a*) separate transmitter and receiver calibration and (*b*) calibration of the ratio of received to transmitted power

from companies that supply that market. For discussion of the relative merits of different passive calibration targets, see Ulaby et al.[86]

The ideal receiver would respond linearly to its input, so that a single calibration at one input level would suffice for all levels. The usual receiver, however, has some nonlinearities due to detector properties and to saturation of its amplifiers by large signals. Figure 16.13 shows a typical input-output curve for a receiver. Two equal increments in input signal (Δ^i), as shown, produce different increments in output because of the nonlinearity of this curve. For this reason, receiver calibration must be performed over a range of input levels, and the nonlinearities must be compensated for in the data processing.

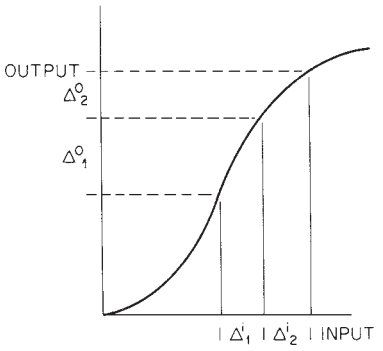


FIGURE 16.13 Typical receiver input-output curve. Illustrated is the effect of nonlinearity.

CW scatterometers depend on antenna beams to discriminate different angles of incidence and different targets. Usually assumptions are made that the antenna pattern has constant gain within the actual 3 dB points and zero gain outside, but this clearly is not an accurate description. If large targets appear in the locations illuminated by the side of the main beam or the minor lobes, their signals may contribute so much to the return that it is significantly changed. Since this changed signal is charged to the direction of the major lobe by the data reduction process, the resulting value for σ^0 is in error. Responses at vertical incidence frequently cause trouble, for vertical-incidence signals are usually fairly strong. Thus the antenna pattern must be accurately known and taken into account in the data analysis. A pattern with strong minor lobes may be simply inadmissible.

The scattering coefficient is determined by applying

$$P_r = \frac{P_t \lambda^2}{(4\pi)^3} \int_{\substack{\text{Illuminated} \\ \text{area}}} \frac{G_t^2 \sigma^0 \, dA}{R^4}$$

The integration is over whatever area is illuminated significantly, including the regions hit by the minor lobes. The usual assumption is that σ^0 is constant over the illuminated area, so that

$$P_r = \frac{P_t \lambda^2 \sigma^0}{(4\pi)^3} \int_{\substack{\text{Illuminated} \\ \text{area}}} \frac{G_t^2 \, dA}{R^4} \tag{16.17}$$

This assumption would be true only if the antenna confined the radiated energy to a very small spread of angles and to a fairly homogeneous region. The resulting expression is

$$\sigma^0 = \frac{(4\pi)^3 P_r}{P_t \lambda^2 \int_{\substack{\text{Illuminated} \\ \text{area}}} \left(G_t^2 / R^4\right) dA} \tag{16.18}$$

Note that only the ratio of transmitted to received power is required, so the technique shown in Figure 16.12*b* is justified. Sometimes R, G_t, or both are assumed constant over the illuminated area, but such an approximation to Eq. 16.18 should be attempted only after checking its validity for a particular problem.

If the result of applying the technique of Eq. 16.18 to a set of measurements indicates that σ^0 probably did vary across the significantly illuminated area, this variation may be used as a first approximation to determine a function $f(\theta)$ describing the θ variation of σ^0, and a next-order approximation then becomes

$$\sigma^0 = \frac{(4\pi)^3 P_r}{P_t \lambda^2 \int_{\substack{\text{Illuminated} \\ \text{area}}} \left[f(\theta) G_t^2 / R^4 \right] dA} \tag{16.19}$$

Proper scattering measurements demand an accurate and complete measurement of antenna gain G_t. This can be a very time-consuming and expensive process, particularly when the antenna is mounted on an aircraft or other metallic object. Nevertheless, complete patterns are a must for good scatter measurements.

Range-Measuring Systems. Radar's ability to separate returns from different ranges can be used advantageously along with directive antenna beams to simplify the scattering measurements. Most ranging scatterometers use either pulse modulation or FM, although more exotic modulations could also be used. The discussion here treats pulse systems, but since all other range-measuring systems can be reduced to equivalent pulse systems, most results are general.

Figure 16.14 shows the way in which pulse measurement of range is used. Figure 16.14*a* shows a circular pencil beam. At angles near grazing, the illuminated patch set by the circular antenna pattern becomes rather long (the patch is an ellipse), and use of the pulse length to confine illumination to a part of the patch is helpful. Indeed, for angles very near grazing, this is the only satisfactory way to resolve small regions. Many systems that use beamwidth to set the measured area near vertical use range resolution for angles beyond, say, 60°.

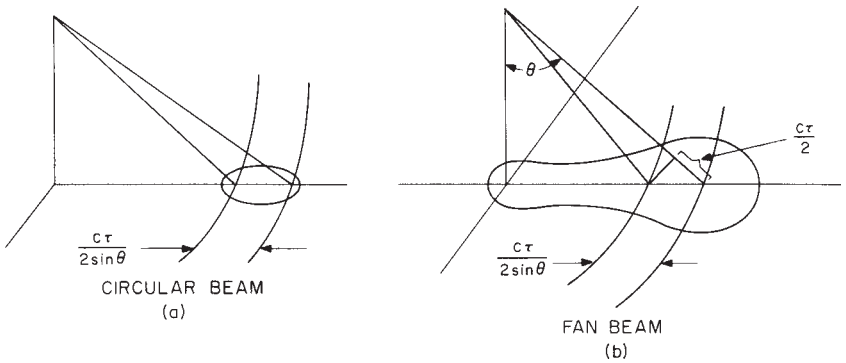


FIGURE 16.14 Range resolution applied to scatterometry: (*a*) improving one dimension of a circular-beam illumination pattern and (*b*) use with a fan beam

Figure 16.14*b* shows an antenna pattern that takes better advantage of the possibilities of range measurement. A fan beam is used to illuminate a narrow strip along the ground, and the range resolution permits separating the returns from different angles by the time they return. This technique is especially effective at angles away from the vertical, for the resolution near the vertical is much poorer than near grazing.

If we assume that σ^0 is essentially constant, the gain is constant, the pulse is rectangular, and the difference in range across a resolution element is negligible, the expression for σ^0 becomes

$$\sigma^0 = \frac{P_r (4\pi)^3 R^3 \sin\theta}{P_t \lambda^2 G_0 \phi_0 r_R} \tag{16.20}$$

where r_R is the short-range resolution.

Janza has reported details of calibration problems with a range-measuring pulsed radar scatterometer.[87,88]

CW-Doppler Scatterometers. A convenient way for an airborne measurement is to measure the scattering coefficient at many angles simultaneously with a CW system in which the relative velocities corresponding to different angles are separated by separating their doppler frequencies. The use of a fan beam with such a system permits the simultaneous measurement of scattering coefficients at points ahead of and behind the aircraft carrying the radar. Figure 16.15 shows this. The pattern of the antenna illumination on the ground is shown intersected by two isodops (lines of constant doppler frequency), with the width of the spectrum between them shown on the diagram. The distance between them can be seen to be

$$\Delta\rho = R(\sin\theta_2 - \sin\theta_1)$$

and

$$\Delta f_d = \frac{2v}{\lambda}(\sin\theta_2 - \sin\theta_1)$$

Thus, the width of the element on the ground is related to the doppler frequency bandwidth by

$$\Delta\rho = \frac{R\lambda}{2v}(\Delta f_d)$$

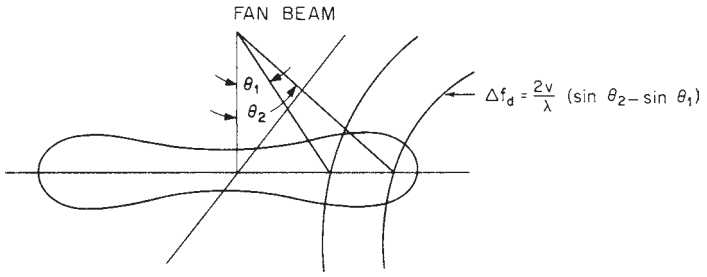


FIGURE 16.15 Resolution in a fan-beam CW-doppler scatterometer

where this technique is applied to the radar equation and the following are assumed:

1. σ^0 constant in the illuminated area
2. Antenna gain constant over its beamwidth Φ and zero elsewhere
3. Range variation across the small illuminated area negligible

$$P_r = \frac{P_t \lambda^2}{(4\pi)^3} \int \frac{G^2 \sigma^0 \, dA}{R^4} = \frac{P_t \lambda^4 \sigma^0 G_0^2 \Phi \Delta f_d}{2vR^2} \tag{16.21}$$

and so

$$\sigma^0 = \frac{P_r}{P_t} \frac{2vR}{\lambda^4 G_0^2 \Phi \Delta f_d} \tag{16.22}$$

Doppler scatterometers need not use fore-and-aft beams. The Seasat[89] and NSCATT[90] spaceborne doppler scatterometers were designed with beams pointed (squinted) ahead and behind the normal to the ground track.

Independent Samples Required for Measurement Accuracy. The Rayleigh distribution describes the fading signal fairly well. If we assume a Rayleigh distribution of fading, the number of independent samples required for a given accuracy is shown in Figure 16.16. The *range* defined in this figure is the range of mean values lying between 5 and 95% of points on the distribution. This accuracy range is independent of any accuracy problems associated with calibration and knowledge of the antenna pattern.

The precision of the measurement depends upon the number of *independent* samples, not on the total number of samples. The number of independent samples can be found from Eq. 16.15 or Eq. 16.16 after suitable analysis. This analysis assumes that only doppler fading contributes to independence but motion from one cell to another also adds independent samples. Thus, the total number of such samples is approximately the product of the number calculated from Eq. 16.13 and the number of ground cells averaged. Figure 16.17 shows some examples of the effect of the angle of incidence on the number of independent samples for a horizontally traveling scatterometer with a forward-pointed beam.

Study of the results obtained in this type of analysis indicates that, in regions where the scattering coefficient does not change rapidly with angle, the widest possible angular width (obtained by a longer pulse or a wider filter for a CW-doppler system) gives the maximum number of independent samples for a given distance traveled along the ground.

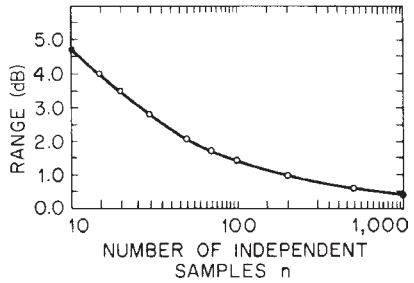


FIGURE 16.16 Accuracy of averages for fading signals

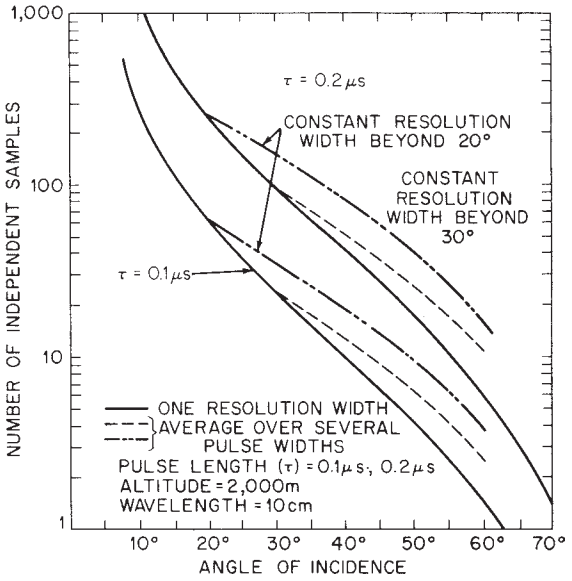


FIGURE 16.17 Examples of the variation with angle of incidence of the number of independent samples for a scatterometer

Near-Vertical Problem. Most published radar return data purporting to include vertical incidence give vertical-incidence scattering coefficients that are too small. This is a consequence of a fundamental problem in measuring near the vertical with a finite beamwidth or pulse length. Near-vertical radar returns from most targets drop off rapidly as the angle with the vertical is increased. Thus, the measuring beamwidth or pulsewidth usually encompasses signals from regions having values for σ^0 many decibels apart. Since the scattering coefficient varies much more rapidly near the vertical than at angles beyond 10 or 20° from the vertical, the problem is much more severe at the vertical. Furthermore, the problem is complicated at the vertical by the fact that the angular scale terminates there, so that a beam centered at the vertical illuminates weaker targets (σ^0) on both sides of its pattern, whereas a beam away from the vertical illuminates stronger signals on one side and weaker signals on the other.

Figure 16.18 shows what happens for a steeply descending curve of σ^0 versus θ. The radar return integral from Eq. 16.1 is a convolution integral; the figure shows the convolution of the beam pattern with the σ^0 curve. Clearly, the average at the vertical is lower than it should be to indicate properly the variation of σ^0 near the vertical.

Figure 16.19 shows an example[91] based on the theoretical scattering coefficient for the sea derived from the spectra reported by the Stereo Wave Observation Project.[92] The effect of different beamwidths is clearly shown.

With a pulse or other range-measuring system, reported values are always in error because, as indicated above, it is almost impossible to resolve a narrow range of angles near the vertical. For short ranges, one can configure the antenna so that a plane wave impinges on the surface. When this is done, the near-vertical scattering coefficient can have its angular variation properly described.[93]

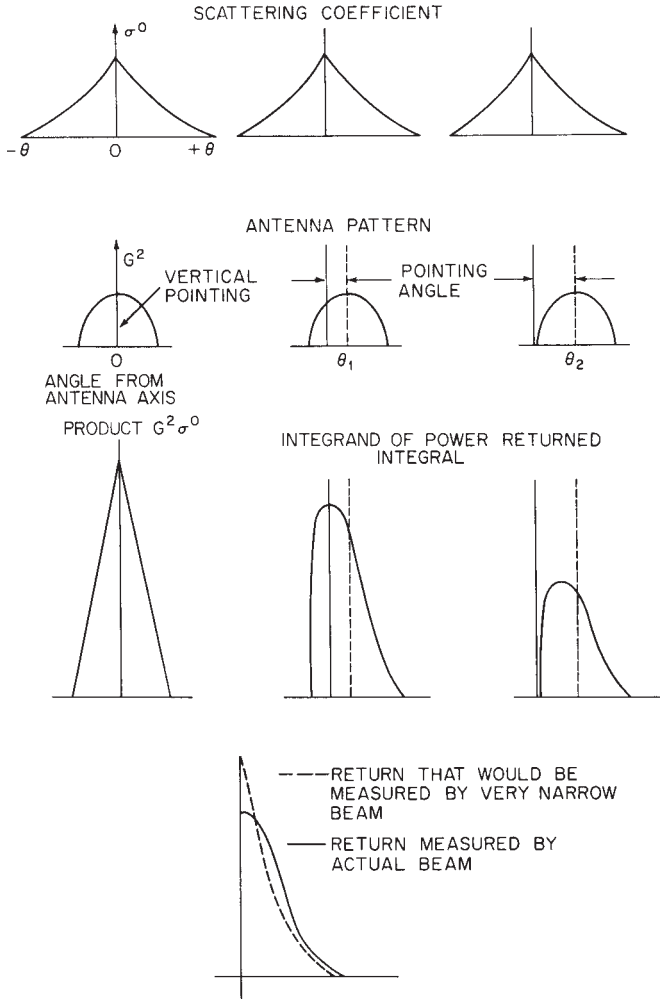


FIGURE 16.18 How finite beamwidth causes a near-vertical error in measuring the scattering coefficient

Ground and Helicopter Scatterometers and Spectrometers. Many ground scattering measurements have been made with systems mounted on boom trucks and helicopters. Most of these are FM-CW systems[94,95] that use wide bandwidth to obtain extra independent samples rather than for fine resolution. Some use very wide bandwidth to obtain fine range resolution to locate sources of scattering.[96] Most have multiple-polarization capability, and some are capable of polarimetry because the phase of two received signals with orthogonal polarization can be measured.[97]

The basic elements of an FM-CW scatterometer are shown in Figure 16.20. The swept oscillator must produce a linear sweep; this is easy with yttrium-iron-garnet (YIG)–tuned oscillators but requires linearizing circuits if tuning uses a varactor.

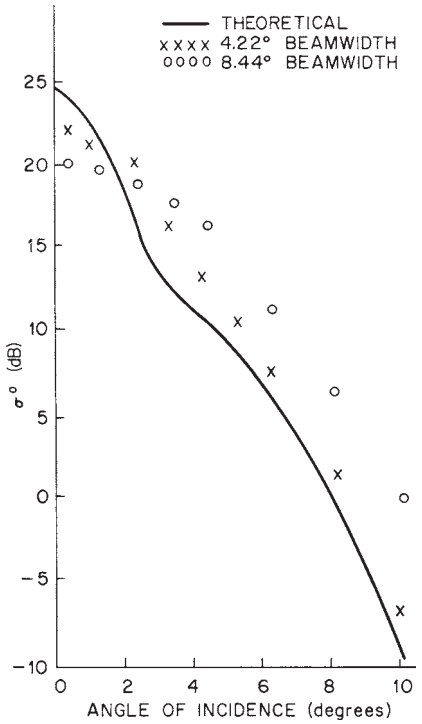


FIGURE 16.19 Effect of antenna beamwidth on the measured scattering coefficient as a function of angle of incidence

Many systems use digital waveform synthesis to obtain the swept waveform. If dual antennas are used (as shown), the overlap of the beams must be considered.[98] Single-antenna systems are sometimes used, with a circulator isolating transmitter and receiver; their performance is somewhat poorer than that of dual-antenna systems because of internal reflections and leakage through the circulator.

Figure 16.21 shows the kind of system that may be used to measure scattering from within a volume. By determining the spectrum of the return, the user can establish the scattering from different ranges. This system has been used in determining the sources of scatter in vegetation[43–45] and snow.

Ultrasonic waves in water can be used to simulate electromagnetic waves in air.[99–101] Because of the difference in velocity of propagation, an acoustic frequency of 1 MHz corresponds with a wavelength of 1.5 mm. Such a wavelength is of a convenient size for many modeling measurements, and, of course, equipment in the 1-MHz region is in many ways easier to operate than equipment in the microwave region; certainly it is much easier to operate and less expensive than microwave equipment operating at a 1.5-mm wavelength.

Acoustic plane waves and electromagnetic plane waves satisfy the same boundary conditions. When the scattering surfaces are not plane and when angles of incidence are rather oblique, the analogy between acoustic and electromagnetic waves is less valid. Of course, acoustic systems cannot simulate cross-polarization.

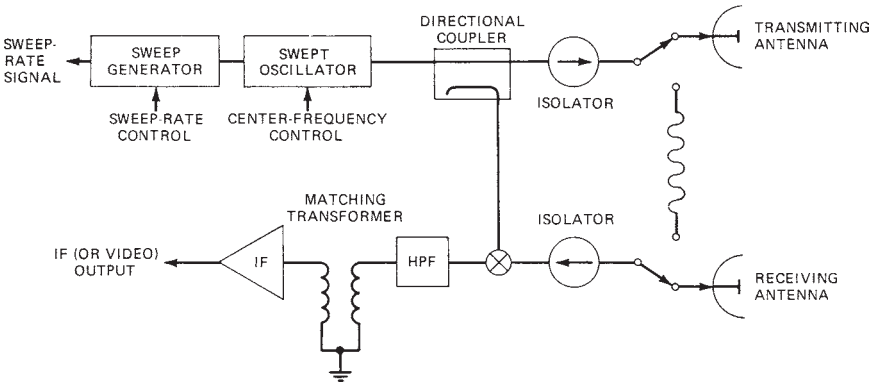


FIGURE 16.20 Basic block diagram of an FM-CW scatterometer RF section

Scattering Coefficients from Images. Radar images produced by real or synthetic aperture radars can be used for scattering coefficient measurement. Unfortunately, most such systems are uncalibrated or poorly calibrated, so the results are somewhat dubious, even on a relative basis, when images are produced on different days. Relative calibration has been introduced into some systems.[20,28,31,102–104] Absolute calibration, which also serves as relative calibration in some cases, can be achieved by using strong reference targets, with the active radar calibrator (ARC) repeaters especially suitable.[85,105] Another approach that has been used is to measure scattering from reference areas with a ground-based or helicopter system that is well calibrated and to compare the images to these measured values.[102,106]

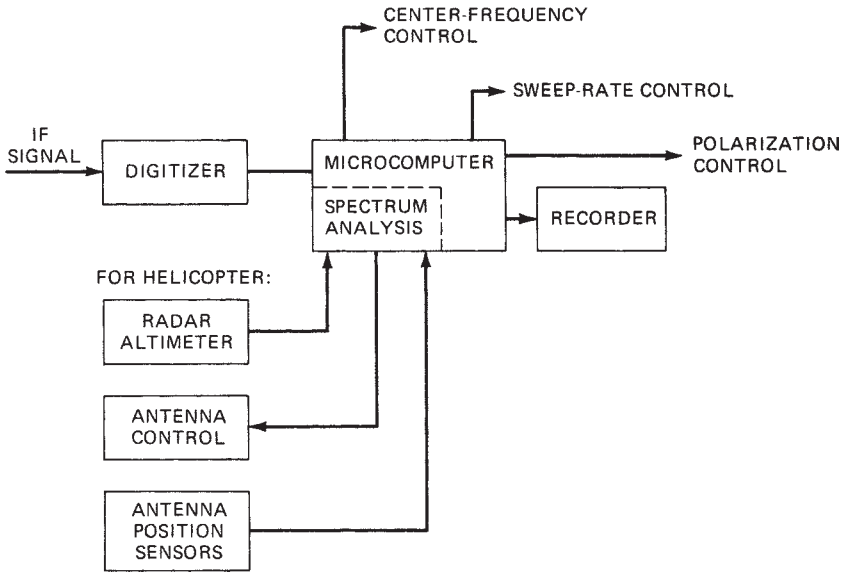


FIGURE 16.21 Basic block diagram of an FM-CW range-discriminating scatterometer: control and data-handling system

Bistatic Measurements. Measurements of ground return when the receiver and transmitter are separated are comparatively rare. These measurements are very difficult to make from aircraft because it is necessary that both transmitter and receiver antennas look at the same ground point at the same time and that the signal be correlated with known antenna look angles. Furthermore, it is difficult to know the polarization, and the exact size and shape of the common area illuminated by the antenna beams are sometimes difficult to determine. For this reason, few bistatic measurements from aircraft have been reported in the literature.[107]

Laboratory bistatic measurements have been made by both the Waterways Experiment Station[42] and The Ohio State University[2,6] groups using electromagnetic waves and by the University of Kansas[100] group using acoustic waves. Bistatic measurements of laser radiation have been made at Bell Telephone Laboratories,[108] and C-band measurements of buildings have been made at the University of Kansas.[109] Other surface-based measurements have also been reported.[110,111]

Bistatic measurements call for complications when made outside the laboratory because an absolute reference for both transmitter power and receiver sensitivity must be used. In the laboratory, however, it is possible to use techniques similar to those for monostatic measurements.

16.6 GENERAL MODELS FOR SCATTERING COEFFICIENT (CLUTTER MODELS)

Scatter measurements made during the 1970s allowed the generation of models for average backscatter from large areas. In particular, these included measurements with the Skylab radiometer-scatterometer RADSCAT[83,112] and with truck-mounted microwave active spectrometers (MAS)[112,113] by the University of Kansas. Two different models were developed based on the same data, one a linear model and one a more complicated formulation. Here we present only the linear model. These models are for *averages*, and the models do not include variations about the average. However, analysis of Shuttle Imaging Radar (SIR) data permits some estimates to be made of the variability to be expected for different sizes of illuminated footprint.

The general characteristics of radar backscatter over the range of angles of incidence have been known for decades. Figure 16.22 shows these. For like-polarized waves, one can break scatter into three angular regimes: near-vertical (the *quasi-specular region*), intermediate angles from 15 to about 80° (the *plateau region*), and near-grazing (the *shadow region*). Cross-polarized scatter does not have separate quasi-specular and plateau regions (the plateau extends to vertical), and too little is known to establish whether a shadow region exists.

For nearly every type of terrain, the measured data fits closely to the form

$$\sigma^0 = A_i e^{-\theta/\theta_i} \tag{16.23a}$$

or

$$\sigma^0{}_{\text{dB}} = 10 \log A_i - 4.3434(\theta/\theta_i) \tag{16.23b}$$

where A_i and θ_i are constants that differ for the near-vertical and midrange regions. Figure 16.23 shows an example of this variation. No theory gives exactly this result,

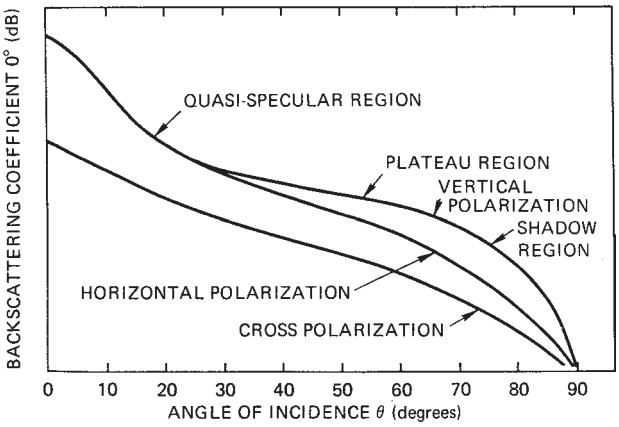


FIGURE 16.22 General characteristics of scattering coefficient variation with angle of incidence (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

but nearly all measurements fit such a model closely, and the model approximates most theoretical curves well over the relevant regions. This simple result means that simple clutter models may be developed and used, although more complex models may be necessary for some remote-sensing applications.

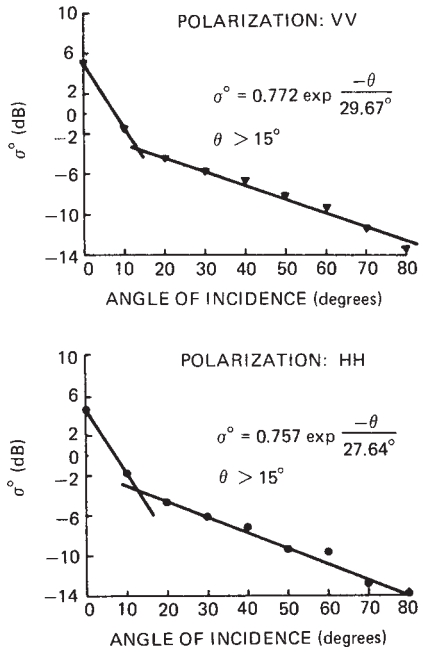

POLARIZATION: VV
σ° = 0.772 exp −θ/29.67°
θ > 15°
σ° (dB)
10 6 2 0 −2 −6 −10 −14
0 10 20 30 40 50 60 70 80
ANGLE OF INCIDENCE (degrees)
POLARIZATION: HH
σ° = 0.757 exp −θ/27.64°
θ > 15°
σ° (dB)
10 6 2 0 −2 −6 −10 −14
0 10 20 30 40 50 60 70 80
ANGLE OF INCIDENCE (degrees)

FIGURE 16.23 Regression of average of all 13.8-GHz cropland data for two years obtained with a microwave active spectrometer (*after R. K. Moore, K. A. Soofi, and S. M. Purduski*[114] *© IEEE 1980*)

The basis for the linear model[114] is a combination of the Skylab results over North America[115] and those from Kansas cropland measurements over three complete seasons with the microwave active spectrometer (MAS).[116] The 13.9-GHz Skylab RADSCAT had a ground footprint of a 10-km circle at vertical to an ellipse of 20 by 30 km at 50°. The MAS had footprints at 50° ranging from 5.5 by 8.5 m at 1.1 GHz to 1.4 by 2.1 m at 17 GHz, but millions of measurements were averaged for the model. Because the Skylab data was at only one frequency and the responses for the two experiments were essentially the same at that frequency, the frequency response shown in the model depends entirely on the MAS measurements.

The summer Skylab observations included deserts, grassland, cropland, and forests, whereas the Kansas measurements were only of cropland. However, early and late in the growing season, the cropland was essentially bare, similar to the summer desert except for soil moisture content. During the height of the growing season, the crops were dense enough so that scatter was similar to that from forests. Thus, the overall model seems representative of summer conditions averaged over all of North America.

The model takes the form

$$\sigma^0_{\mathrm{dB}}(f,\theta) = A + B\theta + Cf + Df\theta \qquad 20° \le \theta \le 70° \tag{16.24a}$$

where A, B, C, and D take on different values for different polarizations above and below 6 GHz. The frequency response below 6 GHz is much more rapid than above 6 GHz. Moreover, at frequencies above 6 GHz the frequency response is independent of angle, so that $D = 0$. For lower frequencies, the frequency response is angle-dependent.

For angles less than 20°, only two points were available, 0° and 10°, so separate frequency regressions were run at each of these angles. The model for these angles is

$$\sigma^0_{\mathrm{dB}}(f,\theta) = M(\theta) + N(\theta)f \qquad \theta = 0°, 10° \tag{16.24b}$$

The frequency responses below 6 GHz differed for the two years, so the models have separate values of the constants for 1975 and 1976. The year 1976 was very dry in Kansas: therefore, the 1975 values are probably more representative, but both are given here. Values of the constants are in Table 16.2. Figure 16.24 shows the clutter model for the midrange of angles as a function of frequency.

TABLE 16.2 Constants for Linear Scattering Model (Summer)*

Eq.	Polarization	Angular Range,°	Frequency Range, GHz	Constant A or M, dB	Angle Slope B or N, dB	Frequency Slope C, dB/GHz	Slope Correction D, dB/(° × GHz)
16.24*a*	*V*	20–60	1–6 (1975)	−14.3	−0.16	1.12	0.0051
	V	20–50	1–6 (1976)	−4.0	−0.35	−0.60	0.036
	V	20–70	6–17	−9.5	−0.13	0.32	0.015
	H	20–60	1–6 (1975)	−15.0	−0.21	1.24	0.040
	H	20–50	1–6 (1976)	−1.4	−0.36	−1.03	
	H	20–70	6–17	−9.1	−0.12	0.25	
16.24*b*	*V* and *H*	0	1–6 (1975)	7.6	. . .	−1.03	
	V and *H*	0	1–6 (1976)	6.4	. . .	−0.73	
	V and *H*	0	6–17	0.9	. . .	0.10	
	V and *H*	10	1–6 (1975)	−9.1	. . .	0.51	
	V and *H*	10	1–6 (1976)	−3.6	. . .	−0.41	
	V and *H*	10	6–17	−6.5	. . .	0.07	

* *After R. K. Moore, K. A. Soofi, and S. M. Purduski*[114] *© IEEE 1980*

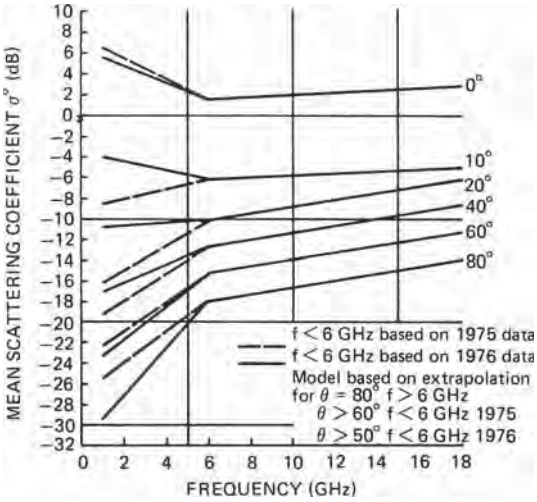


FIGURE 16.24 General land-scattering-clutter model (vertical polarization). Horizontal polarization is very similar. (*after R. K. Moore, K. A. Soofi, and S. M. Purduski*[114] *© IEEE 1980*)

The figure is only for vertical polarization because results are so similar for vertical and horizontal.

Ulaby developed a different, more complex model from the Kansas vegetation data.[117] This model fits curves rather than straight lines to the measured data. For most purposes, the straight-line model is adequate, and it is much easier to use.

A straight-line model for snow-covered grassland similar to that for vegetation depends on a more limited data set.[118,119] The data was for only one season in Colorado when the snow was only about 50 cm deep. This means that the signal probably penetrated to the ground surface at frequencies below about 6 GHz. Nevertheless, the model indicates the kind of results to be expected for this important situation. Table 16.3 gives the resulting constants to use in Eq. 16.24*a*.

TABLE 16.3 Regression Results for Ground-Based Measurements of Snow-Covered Ground*

Time of Day	Polarization	Frequency Range, GHz	Constant *A*, dB	Angle Slope *B*, dB/°	Frequency Slope *C*, dB/GHz	Slope Correction *D*, dB/(° × GHz)
Day	*V*	1–8	−10.0	–0.29	0.052	0.022
Day	*V*	13–17	0.02	–0.37	–0.50	0.021
Day	*H*	1–8	−11.9	–0.25	0.55	0.012
Day	*H*	13–17	–6.6	–0.31	0.0011	0.013
Night	*V*	1–8	−10.0	–0.33	–0.32	0.033
Night	*V*	13–17	−10.9	–0.13	0.70	0.00050
Night	*H*	1–8	−10.5	–0.30	0.20	0.027
Night	*H*	13–17	−16.9	–0.024	1.036	–0.0069

**After R. K. Moore, K. A. Soofi, and S. M. Purduski*[114] *© IEEE 1980*
NOTE: θ = 20 to 70°. Values of coefficients in this table are also considered those of the model.

Snow scatter depends strongly on the free-water content of the upper layer of snow, so scatter is much lower from the wet daytime snow (where solar melting has commenced) than for the dry nighttime snow. Hence, different models must be used for day and night; compare the day and night measurements shown in Figure 16.25. The difference between day and night scatter from snow is even more pronounced at 35 GHz, but the model does not include 35 GHz because no data exists between 17 and 35 GHz.

Although no specific clutter model has been developed for forests, results from the Skylab RADSCAT and Seasat scatterometer show that the Amazon rainforest scatters almost independently of the angle of incidence even near vertical.[120] The mean measured value at 33° was −5.9 ± 0.2 dB at 13.9 GHz. Similar results were found at C band.[121] Observations with SIR-B, SIR-C, and JERS-1 indicated that this lack of angular variation of σ^0 also is present at 1.25 GHz.[122,123]

The models described above are based on averages over very large areas. For this situation, the variability from place to place is small, particularly in the midrange of angles. Figure 16.26 shows the mean and upper and lower decile values measured by the Skylab RADSCAT over North America. The large variation near vertical apparently results from the effect of nearly specular reflection from water bodies. When the footprint is smaller, more variability occurs. This is shown in Figure 16.27 from a study of the variation of scatter observed by SIR-B with averages over different-sized footprints. For small footprints, the scatter varies over a wide range, and system designers must account for this.

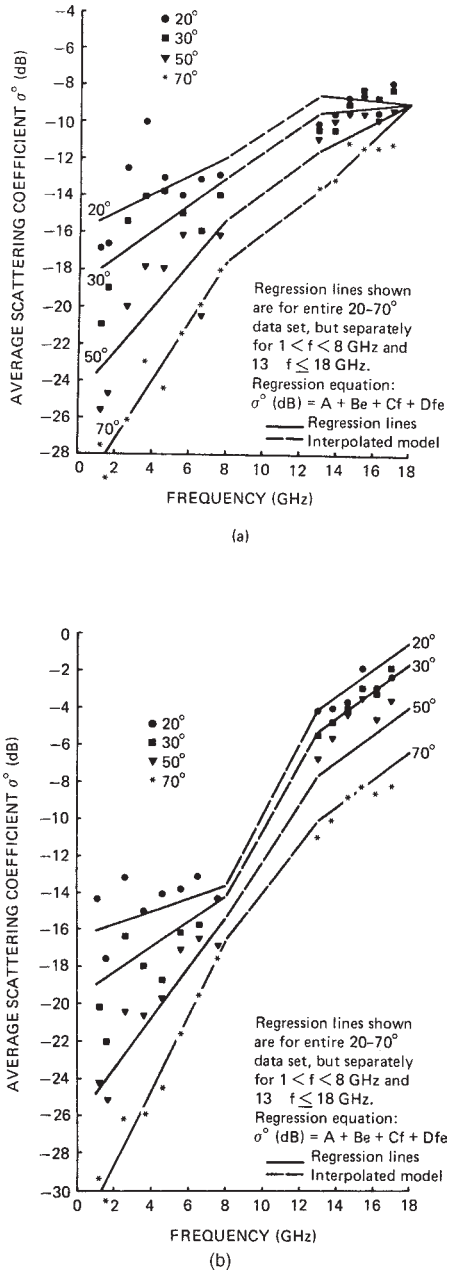

AVERAGE SCATTERING COEFFICIENT σ° (dB)
FREQUENCY (GHz)
20°
30°
50°
70°
Regression lines shown are for entire 20–70° data set, but separately for 1 < f < 8 GHz and 13 f ≤ 18 GHz.
Regression equation:
σ° (dB) = A + Be + Cf + Dfe
Regression lines
Interpolated model
(a)
(b)

FIGURE 16.25 Regressions for vertical-polarization clutter model for snow: (*a*) day and (*b*) night. Note the large differences. Horizontal polarization is similar. (*after R. K. Moore, K. A. Soofi, and S. M. Purduski*[114] *© IEEE 1980*)

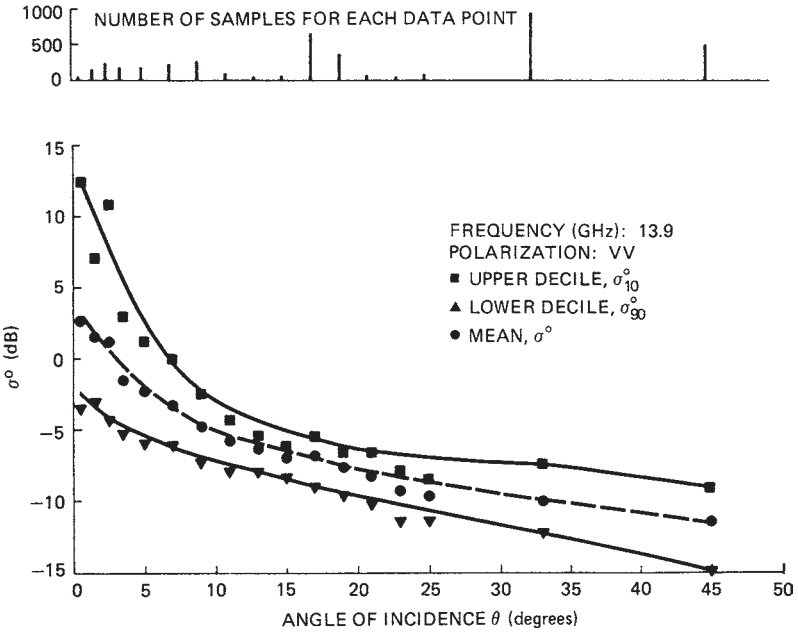


FIGURE 16.26 Angular patterns of the mean, upper decile, and lower decile of Skylab scatterometer observations over North America during the summer season (*from Moore et al., University of Kansas Remote Sensing Laboratory Technical Report 243-12, 1975*)

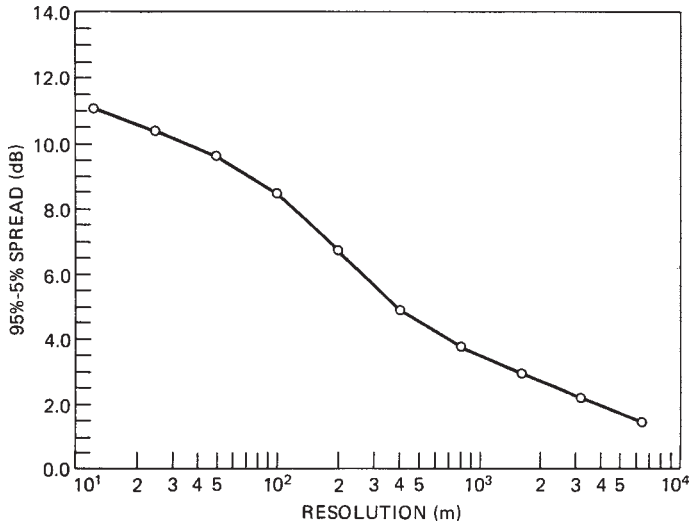


FIGURE 16.27 90% range of pixel amplitude versus resolution

16.7 SCATTERING COEFFICIENT DATA

Numerous programs to gather scattering coefficient data existed prior to 1972, but sizable data collections with accompanying "ground truth" were rare. Since 1972, however, several major programs have changed the situation so that much information is now available. Indeed, this information is so widespread that an adequate summary of the literature is impossible. Hence, this section can only give highlights of the results and major programs. The reader should consult the three major compendia of such data for more information both on results and on bibliography[37,38,40] (note that information is spread through many chapters of these volumes).

Some early scattering-coefficient-measurement programs worth mentioning include those of the Naval Research Laboratory,[23,24] Goodyear Aerospace Corporation,[20] Sandia Corporation (near-vertical data),[124,125] and particularly The Ohio State University.[2,6] From 1972 to 1984, the largest program was at the University of Kansas.[8,9,37,72,83,98,126] Extensive programs were also in France (Centre National d'Etudes Spatiales, Centre National d'Etudes des Télécommunications, Université Paul Sabatier),[12,127] the Netherlands,[10,128] Canada Centre for Remote Sensing (CCRS; especially sea ice),[26,129] and Switzerland and Austria (snow).[130,131] Many of the results from these programs appear in digests of the International Geoscience and Remote Sensing Symposia (IGARSS; IEEE Geoscience and Remote Sensing Society) and journals such as *IEEE Transactions on Geoscience and Remote Sensing* and on *Ocean Engineering, International Journal of Remote Sensing, Remote Sensing of Environment*, and *Photogrammetric Engineering and Remote Sensing.*

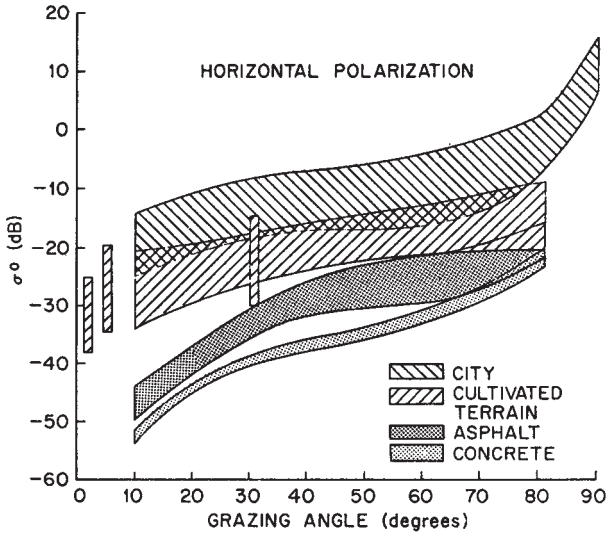
Although calibrations for some of the older data were doubtful, summary presentations are not available for newer data. Accordingly, Figure 16.28 shows an earlier summary based mostly on X-band data. One should be cautious in using this data, but the figure gives a feel for the overall variations. Figure 16.29 is a similar presentation for near-vertical data.[132] Calibration of the systems was good, but the antenna effect discussed in Section 16.5 makes the values from 0 to 5° low.

Effects of Roughness, Moisture Content, and Vegetation Cover. Scattering falls off more rapidly with angles for smooth surfaces than for rough surfaces. Since the roughness that affects radar must be measured in wavelength units, a surface smooth at long wavelengths may be rough at shorter ones. This is illustrated in Figure 16.30,[133] which shows these effects with measurements from plowed fields. At 1.1 GHz, the signal changed 44 dB between 0 and 30° for the smoothest field and only 4 dB for the roughest. At 7.25 GHz the smoothest field was rough enough to reduce the variation to 18 dB.

For most surfaces, cross-polarized scatter is lower than like-polarized scatter, often by about 10 dB. Cross-polarized scatter from smooth surfaces is much less near vertical than elsewhere. Figure 16.31[134] shows this effect. Cross-polarized returns from volume scatterers with elements that are large compared with a wavelength are stronger than for surfaces, sometimes being only 3 dB down.

Scatter depends on the dielectric constant, which depends on moisture content. Thus, scatter from wet soils at angles off vertical is usually much higher than from dry soils. Figure 16.32 shows this.[12] The effect can be many decibels (9 dB in the figure).

Vegetation canopies over soil can contribute to scatter in the various ways shown in Figure 16.33.[135] Figure 16.34[43] shows an example. Most of the scatter from the entire plant came from the top leaves, with enough attenuation there to reduce



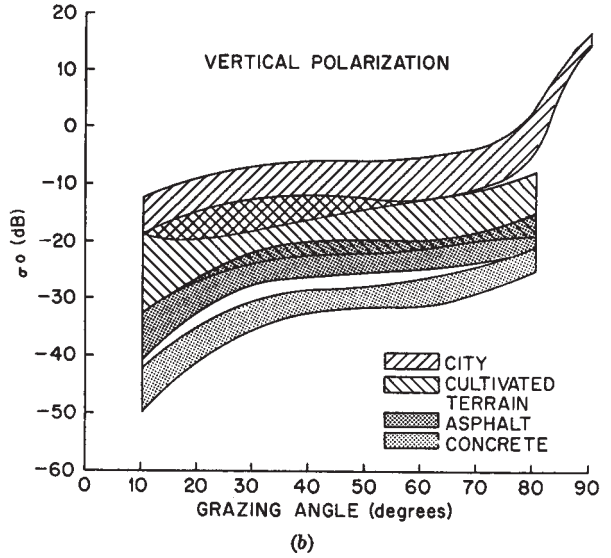


FIGURE 16.28 Boundaries of measured radar data: (*a*) horizontal polarization and (*b*) vertical polarization (*Courtesy of I. Katz*)

the scatter from stem, bottom leaves, and soil to measurable but negligible size. When those leaves were absent, the signals scattered from the soil and lower parts of the plant were about equal to each other and were much larger than when leaves were present.

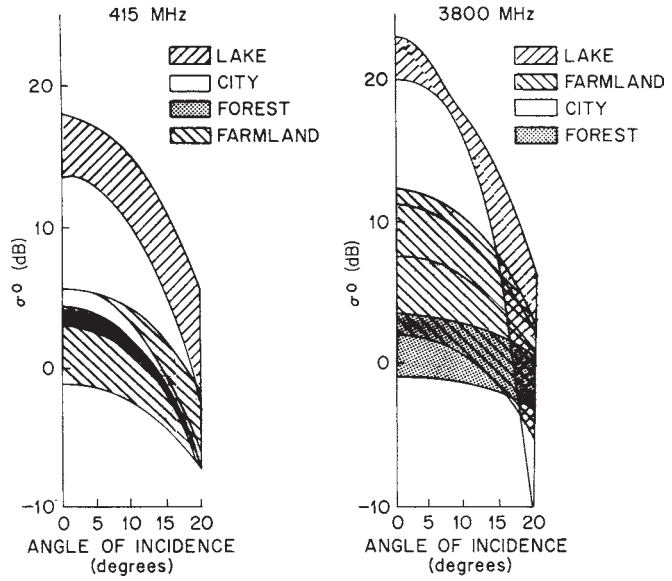


FIGURE 16.29 Boundaries of measured radar return near vertical incidence, based on Sandia Corporation data (*from F. J. Janza, R. K. Moore, and B. D. Warner*[132])

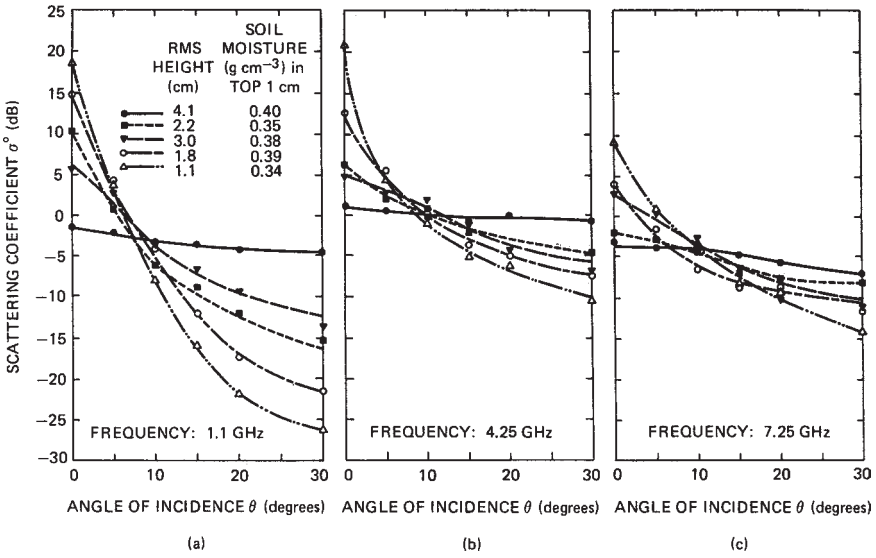


FIGURE 16.30 Angular response of the scattering coefficient for five moist fields with different roughness at (*a*) 1.1 GHz, (*b*) 4.25 GHz, and (*c*) 7.25 GHz (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

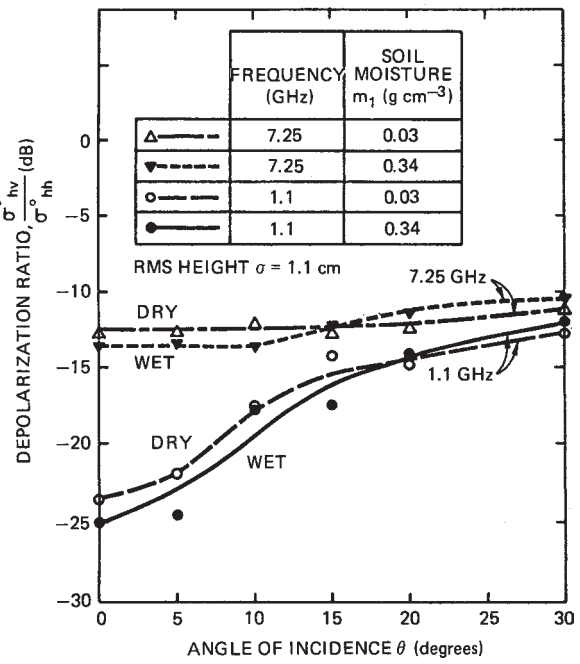


FIGURE 16.31 Angular dependence of the depolarization ratio of a smooth surface (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

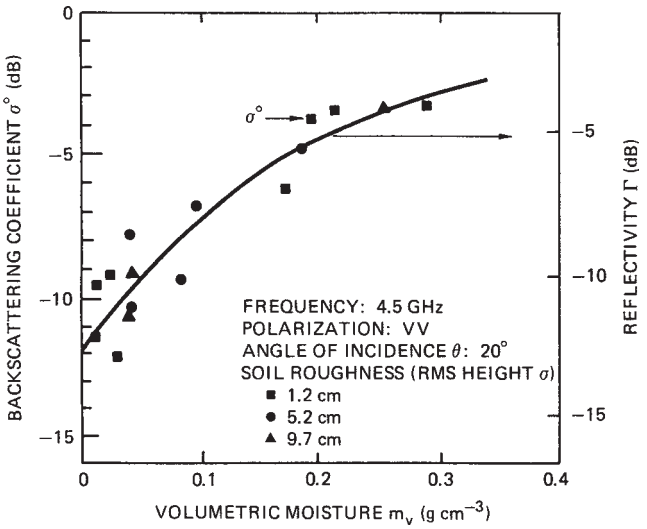


FIGURE 16.32 Measured scattering coefficient σ° (left scale) as a function of soil moisture content for three surface roughnesses. The solid curve is the reflectivity Γ (right scale) calculated on the basis of dielectric measurements. (*after T. LeToan*[12] © *IEEE 1982*)

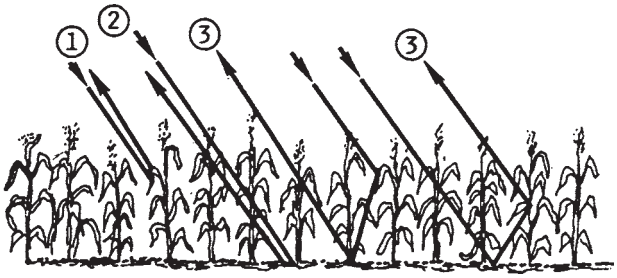


FIGURE 16.33 Contributions to backscatter from a vegetation canopy over a soil surface: (1) direct backscattering from plants, (2) direct backscattering from soil (includes two-way attenuation by canopy), and (3) plant-soil multiple scattering (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

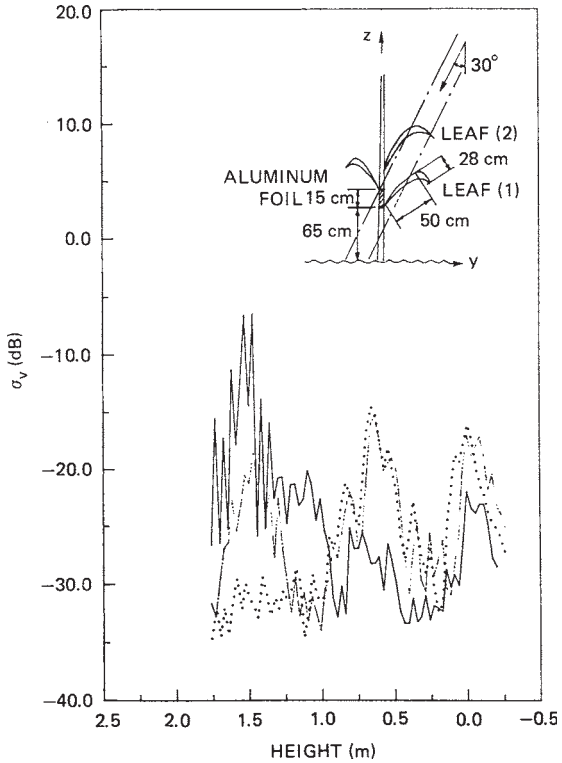


FIGURE 16.34 FM-CW probing scatterometer measurements of a corn plant at 30°. The solid curve is the full plant; the dot-dash curve, leaf 1 removed; the dotted curve, leaf 2 removed. (*after L. K. Wu et al.*[43])

Because volume scatter dominates for dense vegetation, especially trees, σ^0 is nearly independent of the angle of incidence. Figure 16.35[136] shows this with results from X-band imaging of a forest. The figure is a plot of γ rather than σ^0 ($\gamma = \sigma^0/\cos\theta$). At low frequencies such as VHF, this condition changes because the attenuation through the leaves and branches is less.[137]

Soil Moisture. Figure 16.32 shows the size of the effect of soil moisture on σ^0. Soil moisture effects differ for different soils. Dobson and Ulaby[138] showed that this use of moisture expressed in percent of *field capacity* improved the fit between σ^0 and moisture content. Field capacity is a measure of how tightly the soil particles bind the water; the unbound water affects ε more. An empirical expression for field capacity (FC) is[139]

$$FC = 25.1 - 0.21S + 0.22C \text{ percent by weight}$$

where S and C are the percentages (by weight) of sand and clay in the soil. The soil moisture content in terms of field capacity is

$$m_f = 100 m_g / FC \text{ percent}$$

with m_g the percent moisture in the soil by weight. When we use this measure, the relation between σ^0 in dB and m_f is linear even in the presence of moderate vegetation cover, as shown in Figure 16.36[140] The slope of this curve is somewhat different with vegetation cover than it is without, however. Although m_f is apparently at least as good as the volumetric moisture content for relating to σ^0, its use has been questioned.[141]

Soil moisture can affect a radar image, as has been demonstrated in imagery obtained from the Seasat L-band SAR.[142] A simulation experiment[143] showed that one can estimate soil moisture within 20% for 90% of the pixels in an image. Moreover, it showed that resolutions between 100 and 1000 m were superior to finer resolutions

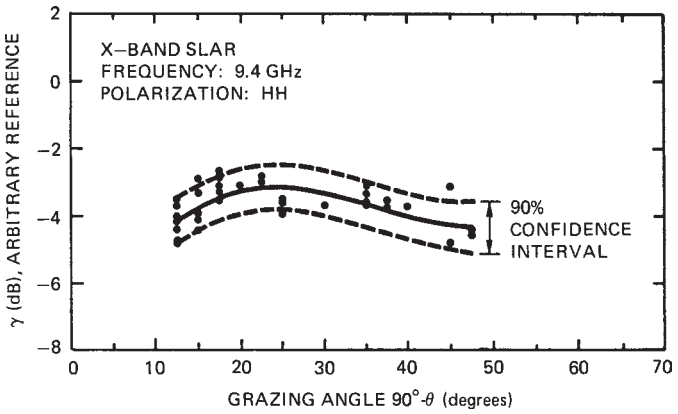


FIGURE 16.35 Measured scattering variation of a forest parcel of old beech trees. Note use of γ (with an arbitrary reference) instead of σ^0 for the ordinate. (*after D. H. Hoekman*[136])

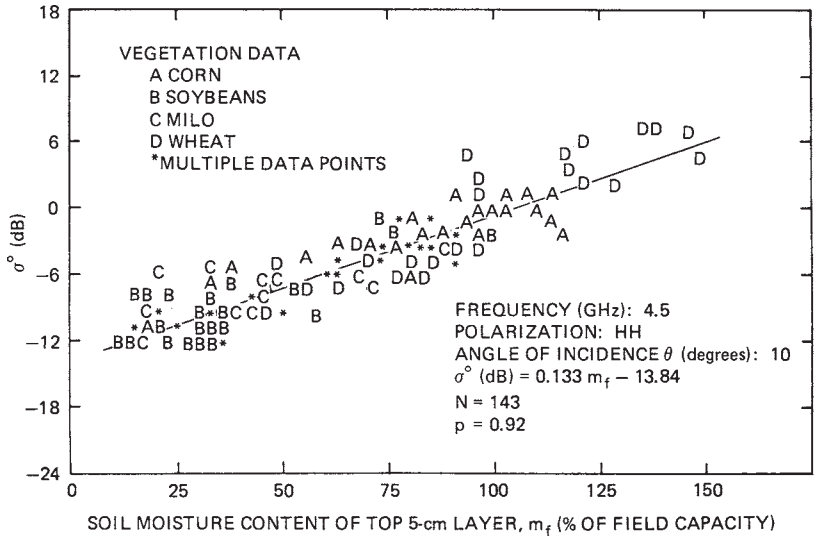


FIGURE 16.36 4.5-GHz scattering coefficient versus soil moisture (percent of field capacity) for vegetation-covered soil (*after F. T. Ulaby et al.*[140])

for this purpose. Most of the spaceborne SARs that followed Seasat have been used in soil-moisture studies.[144,145,146]

Vegetation. Backscatter from vegetation depends on many parameters and varies widely. Thus, although we can develop *average* models like those described in Section 16.6, details are much more complex. The σ^0 varies with season, moisture content, state of growth, and time of day.

Figure 16.37[147] shows the seasonal variation for corn compared with a model presented in the reference. The much larger variation at σ^0 apparently results from the larger effect at vertical of the soil and consequently its moisture content. The rapid 12 dB swing between May 25 and June 1 results from drying of the soil. Even at 50°, where attenuation through the canopy masks the soil effect, the seasonal variation exceeds 8 dB. Diurnal variations are relatively small but finite. They result both from plant moisture changes and from morphological changes (a corn plant actually lifts its leaves "to meet the sun"; morning glories close their flowers at night).

Most crops are planted in rows. This causes an azimuthal variation of σ^0, as shown in Figure 16.38.[148] The modulation shown is the ratio of σ^0 looking parallel to the rows (more vegetation) to that looking normal. This phenomenon is much more pronounced at the lower frequencies.

Some general properties of vegetation scatter are visible in Figure 16.39.[149] At low frequencies, the decay with θ is rapid out to about 20° and then more gradual; most of the steep part results from surface echo. At higher frequencies, the plant attenuation prevents a significant surface echo, so the angular variation is more uniform.

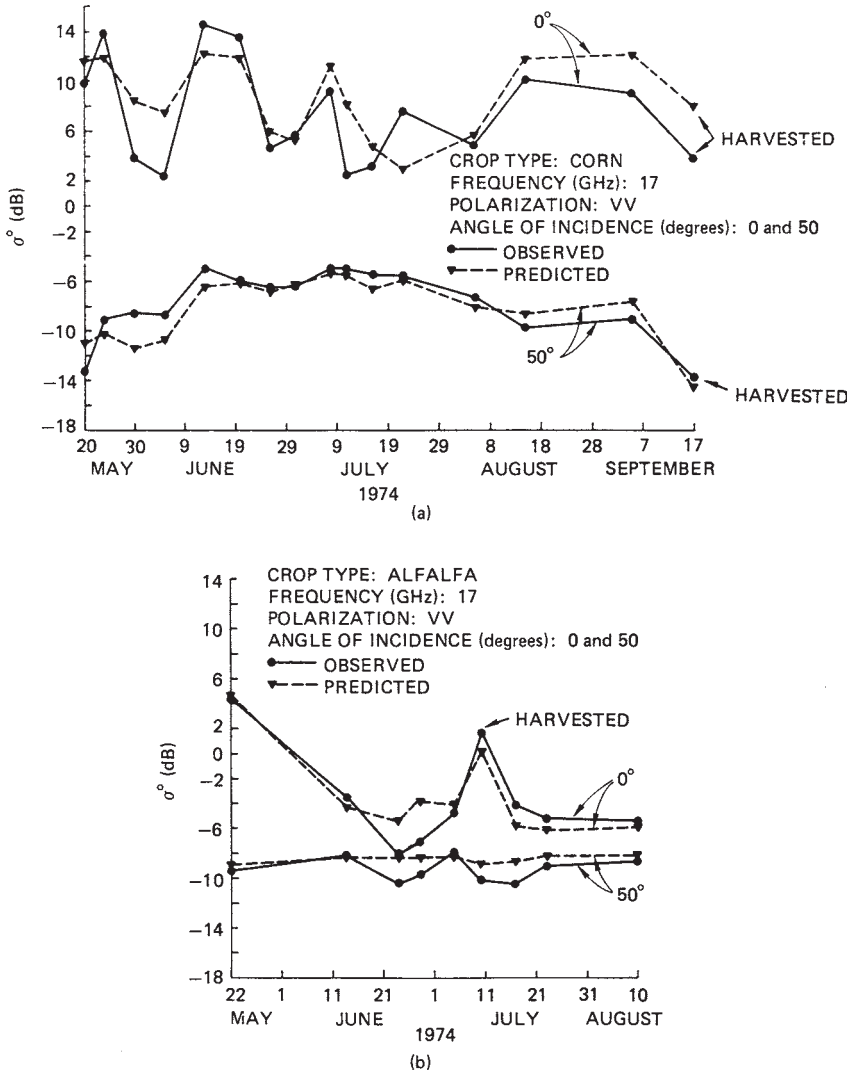


FIGURE 16.37 Time variation of scatter from (*a*) corn and (*b*) alfalfa at incidence angles of vertical and 50° (*after E. Attema and F. T. Ulaby*[147])

Cross-polarized signals at vertical are negligible, so even at low frequencies, the cross-polarized σ^0 varies uniformly. At both high and low frequencies, it is about 10 dB below the like-polarized σ^0.

Snow. When snow covers the ground, much of the scatter is from the snow rather than the underlying ground. Snow is both a volume-scattering and an attenuating medium. When the snow is dry, scatter comes from a large volume; when it is wet, the scattering volume is much less because of higher attenuation. As a result,

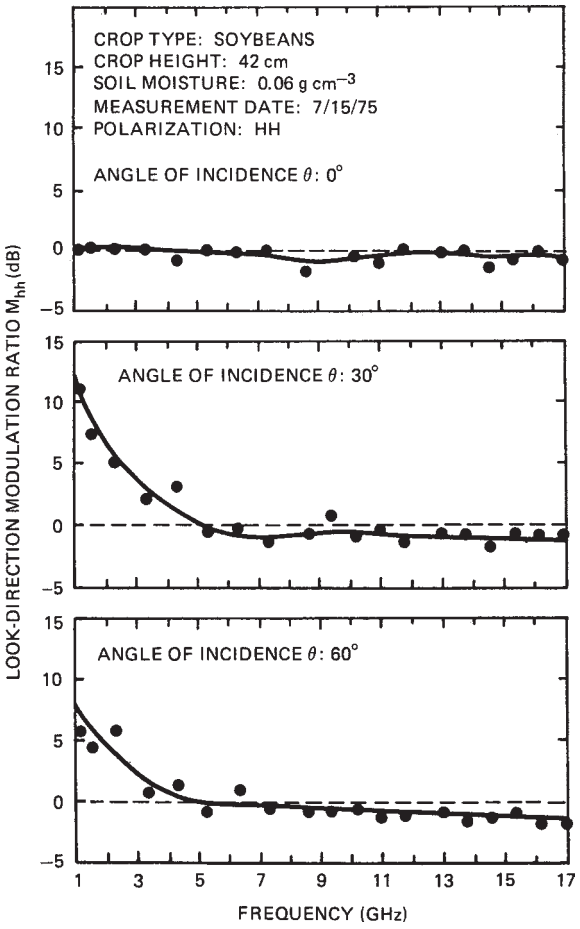


FIGURE 16.38 Frequency response of the look-direction modulation ratio for a soybean field with horizontal polarization at incidence angles of 0, 30, and 60° (*after F. T. Ulaby, R. K. Moore, and A. K. Fung*[37])

σ^0 decreases rapidly as the sun melts the top layer. Figure 16.40[150] illustrates how fast this can be and also shows that the effect is much greater at the higher frequencies where attenuation is greater. Figure 16.41[151] shows the angular variation seen for snow-covered ground. Off-vertical scattering is much greater at higher frequencies. For the 58-cm depth shown, much of the scatter at 1.6 and 2.5 GHz is probably from the underlying surface.

Some reports state that there are *radar hot* spots in snow cover, particularly at 35 GHz. These reports result from improper interpretation of variations that are due to normal Rayleigh fading of the signal. Scatter from snow comes from many centers within the illuminated volume, so the conditions for Rayleigh fading are met.

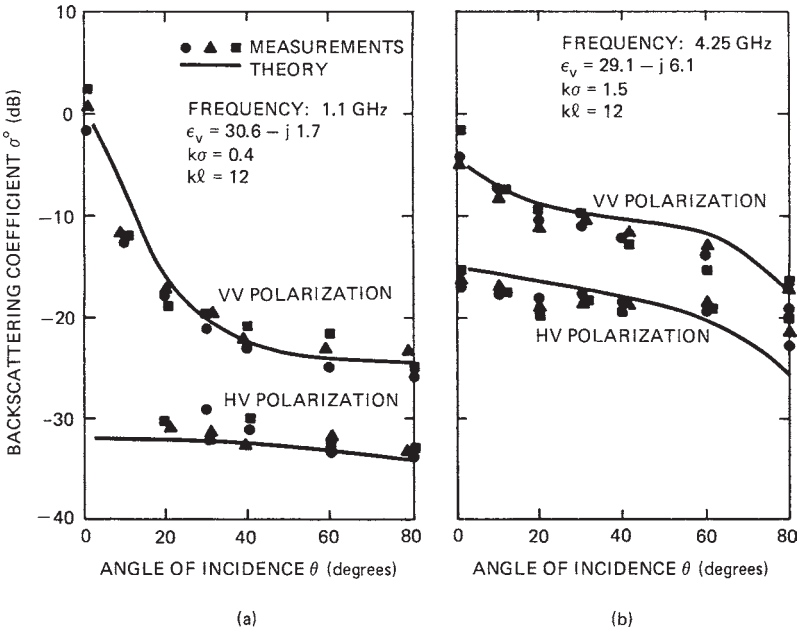


FIGURE 16.39 Comparison of model calculations with measurements at (*a*) 1.1 GHz and (*b*) 4.25 GHz (*after H. Eom and A. K. Fung*[149])

Measurements with suitable averaging in frequency or illumination angle demonstrate that snow-covered surfaces scatter essentially uniformly except for the effects of the multipath fading.

Sea Ice. Sea ice is a very complex medium. Ice observers characterize it in many different categories that depend on thickness, age, and history of formation.[152] Hence, one cannot characterize its radar return in any simple way; in this sense, it is like vegetation. The most important ice types from a radar point of view are first-year (FY 1 to 2 m thick), multiyear (MY > 2 m thick), and a conglomeration of thinner types (< 1 m thick).

Like snow, sea ice influenced by solar melting and above freezing temperatures scatters microwaves very differently from the more normal cold-surface ice. In winter, the cold MY ice scatters much more than cold FY ice. In summer, σ^0 for MY ice decreases to about the same level as that of FY ice. Figure 16.42[153] shows this and typical angular responses. These curves are for 13.3 GHz, but the results would be similar at any frequency down to S band. Figure 16.43[126] shows the frequency variation of σ^0 for various kinds of ice. Shore-fast ice is grounded to the bottom at the shoreline; in this case, it is probably MY. Gray ice is one of the types thinner than FY.

Kim[72] developed a theory that explains a wide range of sea-ice σ^0 measurements. From this and extensive data from the literature on ice properties, Figure 16.44[91] shows the ranges of FY and MY scattering under winter conditions. Clearly higher frequencies are better for identifying ice types than lower frequencies, and

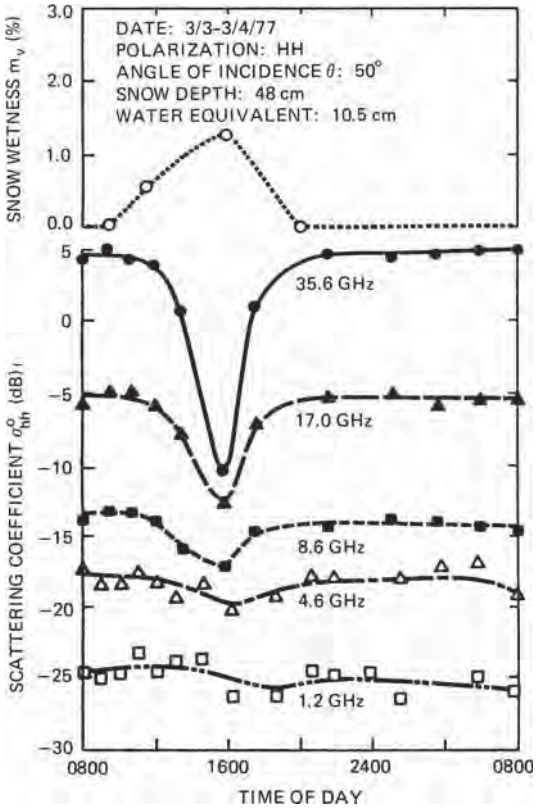


FIGURE 16.40 Diurnal patterns of σ^0 and liquid-water content for snow at several frequencies. Note the extreme variation of the K_a band as the sun starts to melt the surface. (*after W. H. Stiles and F. T. Ulaby*[150])

discrimination is not possible below about 5 GHz. At L band and below, the differences between MY and FY ice are small even in winter. This means that imaging radars can easily distinguish ice types by intensity alone at the higher frequencies in winter but not in summer. This fact is the basis for operational ice-monitoring systems by the Soviet Union [using the Toros K_u-band side-looking airborne radar (SLAR)][154] and Canada[155] (using a modified X-band APS-94 SLAR and the STAR-1 X-band SAR). A prime motivation for the Canadian Radarsat SAR was monitoring of sea ice, which the system has been doing successfully since 1995.[156,157] The Russian X-band real-aperture radars in the Okean series have been used for similar purposes.[158–160]

Snow cover on ice can mask ice scatter itself as with snow on land. Since the arctic is relatively dry, most areas have little snow, but snow does make distinguishing ice types difficult at times. This is particularly true in the Antarctic, where snow is more prevalent on the sea ice.[161]

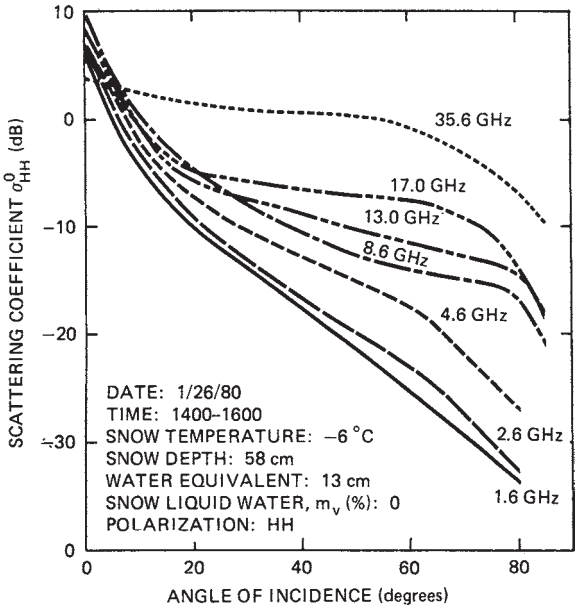


FIGURE 16.41 Angular response of σ^0 of dry snow at different frequencies. Rapid falloff at lower frequencies apparently results from penetration to the smooth ground surface. (*after W. H. Stiles et al.*[151])

Programs to learn microwave properties of sea ice have been numerous because of the importance of arctic operations and meteorology. Microwave remote sensing is necessary to monitor ice properties in the arctic owing to the long winter night, frequent cloud cover, and inaccessibility.

16.8 POLARIMETRY

Several synthetic-aperture imaging radars are capable of measuring the full complex polarization matrix. Probably the first of these was an airborne system built by NASA's Jet Propulsion Laboratory. The first one in space was the Shuttle Imaging Radar–C (SIR-C). Although use of multiple polarizations dates from the early days of imaging radars, the measurement of phase between the received signal with different polarizations is more recent, dating from the late 1980s. Few, if any, full-polarimetric data sets exist of the type described above for single polarizations. For more complete discussions of radar polarimetry, consult references Ulaby and Elachi,[162] Sletten and McLaughlin,[163] and van Zyl and Kim.[164]

Since polarimetric radars use defined phases for both transmit and receive, the signals must be described in the form used for elliptical polarization. This is illustrated in Figure 16.45.

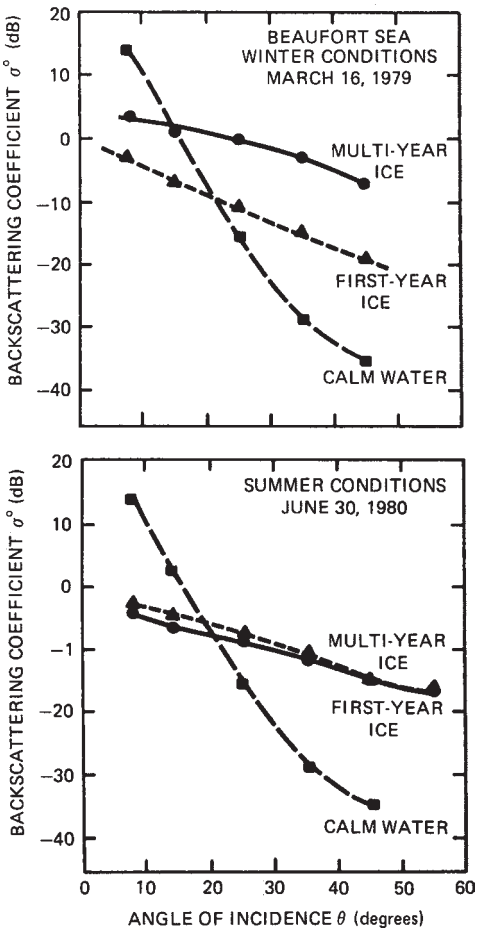


FIGURE 16.42 Comparison of sea-ice scattering at 13.9 GHz in summer and winter (*after A. L. Gray et al.*[153] © *IEEE 1982*)

When $\chi = 0$, the polarization is linear with the E vector in the direction given by ψ. When $\chi = \pm 45°$, the polarization is circular, with +45° for left-hand and −45° for right-hand. When $0 < |\chi| < 45°$, the polarization is elliptical.

Analytically, the electric field may be described as

$$\mathbf{E} = E_h \mathbf{1}_h + E_v \mathbf{1}_v \tag{16.25}$$

where $\mathbf{1}_h$ and $\mathbf{1}_v$ are unit vectors in the h and v directions. The instantaneous fields are given by Eq. 16.26, where the δs show the different phases for the components of **E** and k is the wavenumber.

$$e_h(t) = \mathrm{Re}\, E_h e^{j(\omega t - kx + \delta_h)} \tag{16.26a}$$

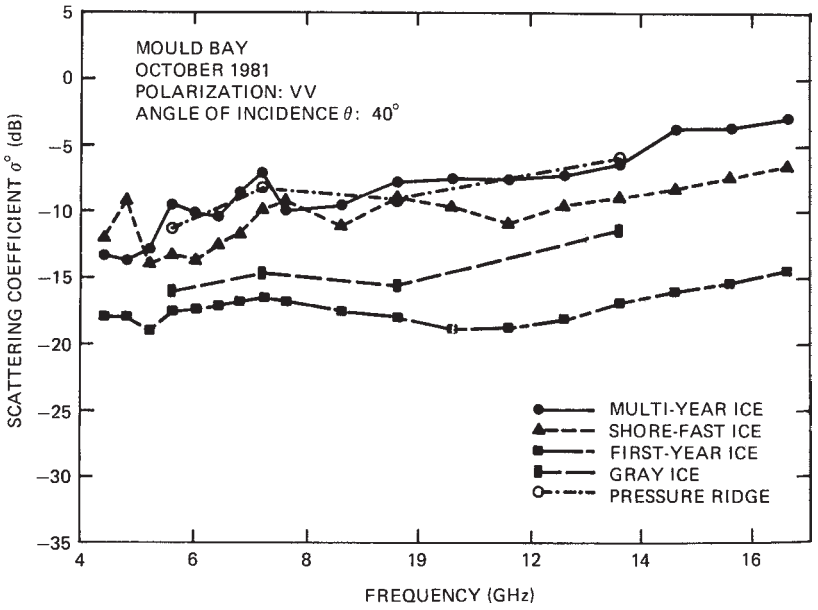


FIGURE 16.43 Example of frequency response of σ^0 for different kinds of sea ice (*from Y. S. Kim*[72])

and

$$e_v(t) = \mathrm{Re}\, E_v^{\,j(\omega t - kz + \delta_v)} \tag{16.26b}$$

In complex format

$$E_h = E_{h0} e^{j\delta_h} \quad \text{and} \quad E_v = E_{v0} e^{j\delta_v}$$

If we take $\delta = \delta_v - \delta_h$ and set $\delta_v = 0$ as a reference, we can write

$$E = E_{h0} e^{-j\delta} + E_{v0} 1_v$$

Thus, for a single wave, we need only have three independent parameters. In radar, we must consider both transmitted and received polarimetric signals, so the need for four magnitudes and two phases.

Another way to describe polarimetric signals is to use the matrix of Stokes parameters:

$$\boldsymbol{F} = \begin{bmatrix} I_0 \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} |E_h^{\,2}| + |E_v|^2 \\ |E_h^{\,2}| - |E_v|^2 \\ 2\,\mathrm{Re}(E_h E_v^*) \\ 2\,\mathrm{Im}(E_h E_v) \end{bmatrix} \tag{16.27}$$

The individual Stokes parameters I_0, Q, U, and V are defined as shown in Eq. 16.27.

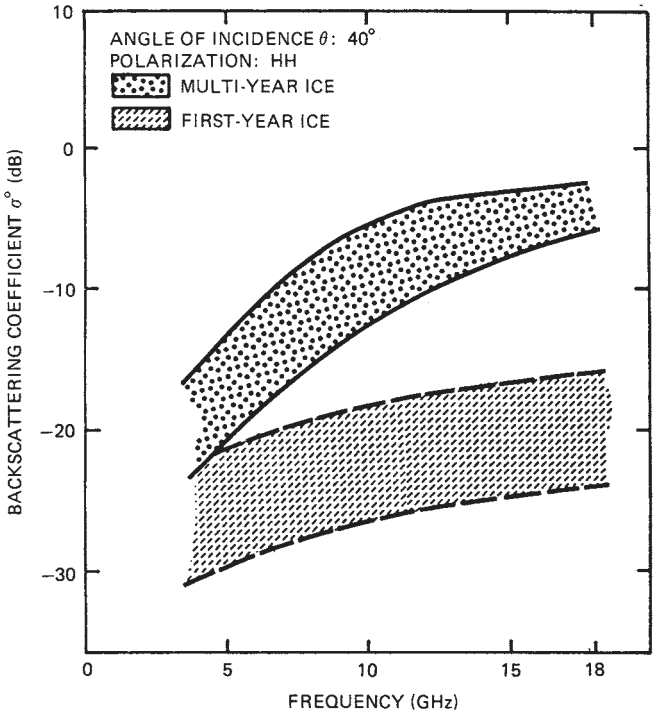


FIGURE 16.44 Measurement-based theoretical σ^0 variations for first-year and multiyear sea ice. Ranges are determined by using known variations of ice characteristics. (*after Y. S. Kim et al.*[126])

Some of the return signals from a resolution cell maintain their polarization characteristics over time and space, whereas others have a random polarization. This occurs, as for sunlight, when the polarization ellipse changes its properties randomly and rapidly with time or with small differences in angle. When both the persistent and random parts

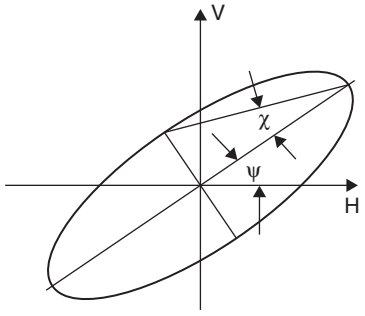


FIGURE 16.45 Polarimetric ellipse: χ is the ellipticity angle, and ψ is the orientation angle. The ellipse is the locus of the end of the E vector throughout a cycle.

are present, the target is said to be partially polarized; when no random component is present, the target is fully polarized. Radar signals are frequently only partially polarized, especially where multiple bounces occur within the target area.

For the nonrandom part, we must define a Stokes vector using an ensemble average of each component; the averaging may be over time or look angle. Thus we have

$$\boldsymbol{F} = \begin{bmatrix} \left\langle |E_h|^2 + |E_v{}^2| \right\rangle \\ \left\langle |E_h|^2 - |E_v|^2 \right\rangle \\ \left\langle 2\mathrm{Re}(E_h E_v) \right\rangle \\ \left\langle 2\mathrm{Im}(E_h E_v) \right\rangle \end{bmatrix} \tag{16.28}$$

when the wave is completely polarized,

$$I_0^2 = Q^2 + U^2 + V^2 \tag{16.29}$$

but when it is partially polarized,

$$I_0^2 > Q^2 + U^2 + V^2 \tag{16.30}$$

In fact, when the wave is completely unpolarized (sunlight, for example), E_v and E_h are uncorrelated, so U and V are both zero.

Waves of this type are used in polarimetric radar; however, to see how the scattering coefficient works, we need to consider both the incident and scattered wave.

At this point, we must introduce the scattering matrix $\bar{S}$. The received field may be represented by

$$\mathbf{E}^r = \frac{e^{-\mathrm{jkR}}}{R} \bar{S}\, \mathbf{E}^t \tag{16.31}$$

where

$$\mathbf{E}^r = \begin{bmatrix} E_v^r \\ E_h^r \end{bmatrix} \quad \text{and} \quad \mathbf{E}^t = \begin{bmatrix} E_v^t \\ E_h^t \end{bmatrix}$$

and

$$\bar{S} = \begin{bmatrix} S_{\mathrm{vv}} & S_{\mathrm{vh}} \\ S_{\mathrm{hv}} & S_{\mathrm{hh}} \end{bmatrix} \tag{16.32}$$

In the usual reciprocal media, $S_{\mathrm{vh}} = S_{\mathrm{hv}}$. Since the choice of a phase reference is arbitrary, there are then three independent magnitudes ($|S_{\mathrm{vv}}|$, $|S_{\mathrm{hh}}|$, $|S_{\mathrm{hv}}|$), but only two independent phases ($\angle S_{\mathrm{hh}}, \angle S_{\mathrm{hv}}$). These quantities can be used to describe the properties of the polarized part of the echo from a target.

We can also describe the scattering using the Mueller matrix that is related to the Stokes matrix. The reader is referred to the literature for further descriptions of the Mueller matrix.[162,165]

The usual way to obtain polarized responses is to transmit alternate vertically and horizontally polarized pulses. Presuming there is almost no change in the target during the interpulse interval, the responses may be combined to produce the scattering matrix or Mueller matrix. By combining the signals in processing, this method permits synthesizing *equivalent transmitted polarizations* with any ellipticity and orientation.

A commonly used way to describe the polarization characteristics of a target is the *polarization signature*.[32] This consists of two three-dimensional graphs. For the first graph, called copolarized, one uses the components of the received signal that are *the same* as the transmitted-signal polarization. The second graph uses components of the received signal that are *orthogonal* to the transmitted signal.

Figure 16.46 gives a widely quoted example of this type of display for images of San Francisco.[166] The axes in the horizontal plane are the orientation angle for the synthesized transmitted signal ψ_t and its ellipticity angle χ_t. The vertical axis is relative power. Values of $\psi_t = 0°$ and 180° are horizontal polarization, whereas $\psi_t = 90°$ is vertical polarization. Linear polarization occurs when $\chi_t = 0°$, and right and left circular polarizations occur when $\chi_t = \pm 45°$. When the minimum is above zero, the pedestal beneath it corresponds to the unpolarized signal.

The ocean image in Figure 16.46*a* shows that the polarization is essentially linear, with the VV signal stronger than the HH. The cross-polarized response shows essentially no cross-polarized signal for linear transmission, but some for circularly polarized transmission.

For the park shown in Figure 16.46*b*, the vertically polarized linear signal is slightly higher than the horizontally polarized one. There is some cross-polarized response for the linear signal and some unpolarized signal. For the urban area shown, the strongest responses are tilted in orientation in both the copolarized and cross-polarized cases.

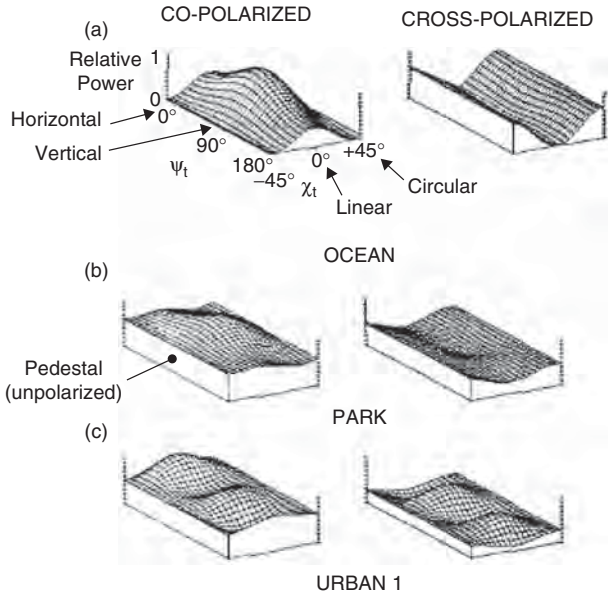


FIGURE 16.46 Selection of polarization signatures from SAR image of San Francisco: (*a*) an ocean area, (*b*) a large park, and (*c*) one of several urban areas (*after D. L. Evans*[32] *© IEEE 1988*)

The actual numbers for the copolarized response for this case are $\chi = 0°$ (VV) and $\psi = 20°$.[32] The reference presents similar plots showing separately the polarized and unpolarized responses.

Because of the complexity of this representation and others for polarimetric signals, one cannot as readily provide curves of response as for single-polarization images. Hence, we do not find many catalogs of polarimetric scattering responses.

Nevertheless, many authors have described the use of polarimetric images. Some of the papers using the term *polarimetric* really refer to the use of just HH, VV, and HV polarizations without regard for the phase. In this sense, they are using these images in the way that like- and cross-polarized images have been used since the start of imaging radars.[167] Others, however, take full advantage of the full polarization matrix.

One way to use the full matrix is to synthesize polarizations that either emphasize or suppress particular classes of targets. For example, in Figure 16.46*c*, one could synthesize a 20° orientation angle for a linear polarization to emphasize this class or use vertical polarization to suppress the dominant class in the image. Various authors[168,169] have shown that one can synthesize an elliptical polarization that increases the target-to-clutter ratio where the target is, for example, a manmade object. Swartz et al.[169] found a polarization that gave a 9.4 dB target-to-clutter ratio where the target was an urban area and the clutter was represented by a park in the San Francisco image used to produce Figure 16.46. He found this was obtained with transmitter polarization having $(\psi_e, \chi_t) = (-41.3°, -6.4°)$ and receiver polarization of $(\psi_e, \chi_t) = (60.3°, 3.5°)$. This compares with the best result using like- and cross-polarizations without phase coherence of 7.3 dB.

Others use the three independent magnitudes of the scattering matrix and the phase angle δ between the HH and VV responses. Research has shown there is little useful information in the phase angle relating HH to cross polarization. A common use of these data is in producing state vectors to employ in discrimination of target areas, with the components of the vectors being the three magnitudes and the phase angles for each frequency used. These vectors are then used in various statistical algorithms to identify different target classes.[170,171,172] This approach has also been used in other forests,[173,174] agricultural areas,[175,176] sea ice,[177] and snow,[178] and to identify surface classes of geologic significance.

An example of the use of phase difference in discriminating surface classes in the Amazon basin is illustrated in Figure 16.47.[179] Note the significant differences between C and X band for Macrophyte and Flooded forest. These differences could be used as discriminators to identify these classes, but normally they would simply be additional elements in the state vectors used in statistical algorithms.

16.9 SCATTERING COEFFICIENT DATA NEAR GRAZING

Conditions for backscatter near grazing incidence are sufficiently different from those at steeper incidence that they must be described separately. Here, we draw upon the work of Billingsley.[17] They collected data over a wide range of terrain and unlike most previous near-grazing measurements, also collected extensive "ground-truth" information and accurately calibrated their radars. Moreover, these data were over a wide range of frequencies: VHF (167 MHz), UHF (435 MHz), L band (1.23 GHz), S band (3.24 GHz), and X band (9.2 GHz). They had 43 different target areas in different parts of the U.S. and western Canada.

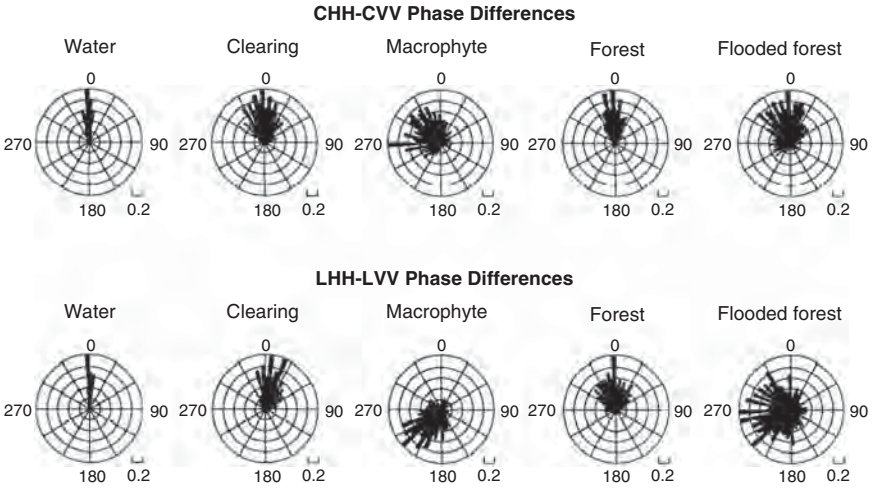


FIGURE 16.47 Phase differences for different surface classes in the Amazon basin at C band and L band[179]

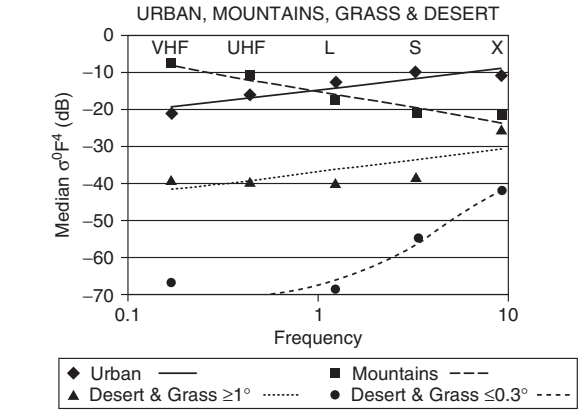
The results were described in terms of $\sigma^0 F^4$, which was called *clutter strength*. F is a propagation factor that accounts for multipath, attenuation, etc., but could not be measured separately.

Whereas *angle of incidence* is used to describe results in studies where pointing is nearer vertical, depression or grazing angle is a more appropriate description when dealing with angles where incidence is closer to horizontal. Grazing angle is the complement of incidence angle. The measurements were all at low *depression* angles, with this term used in lieu of *grazing* angle because it can be defined in terms of antenna pointing, whereas grazing angle also depends on local slope, which is both variable and, in general, unknown.

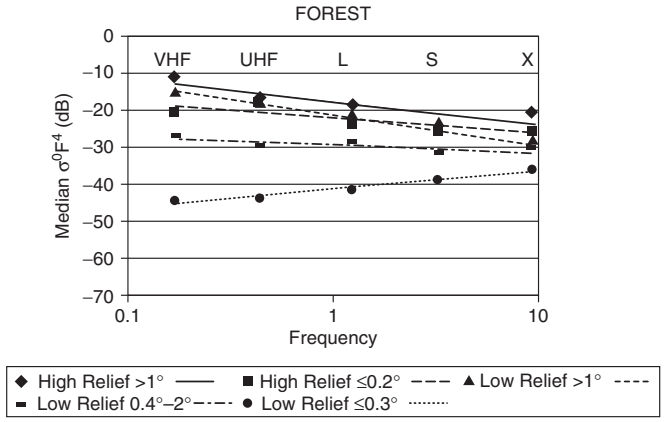
Images taken at near-grazing angles tend to be extremely "patchy" because any projections above the surface (trees, hills, buildings, fences, power lines, machinery, and vehicles) present surfaces more nearly normal to the beam. Thus nearby pixels may have clutter strength differing by tens of dB. Moreover, small slopes facing the radar increase the grazing angle, resulting in stronger echoes, whereas slopes away from the radar decrease the signal or may obliterate it by shadowing.

Because of this effect, the probability distributions of near-grazing echoes are very different from those at more moderate incidence angles. Although some small regions without significant projections or forward slopes may have Rayleigh distributions or (if only one large scatterer is present) Ricean distributions, most areas have other distributions—often Weibull or even log-normal. The result is that the mean estimated $\sigma^0 F^4$ is often much higher than the median; a few targets 10 or 20 dB higher can raise the mean a lot; even though, the targets occupy only a tiny fraction of the area. Thus, one should be careful in using mean values for radar design; the median values are more representative.

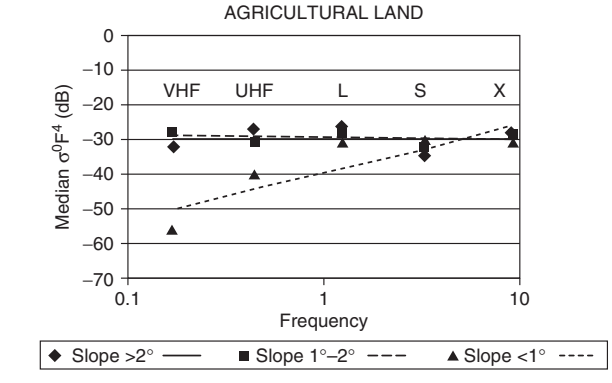
Billingsley[17] presented his results both in terms of mean values and median values. Here, we report only the median values because they tend to be more meaningful for radar design than the values distorted by occasional strong targets. For most areas, little difference was found between vertical and horizontal polarizations, so the results were reported for data groups including both polarizations. Figure 16.48 shows the results, grouped by target class.



(a)



(b)



(c)

FIGURE 16.48 Median values of $\sigma^0 F^4$ at low depression angles (based on Table 3.6 of Billingsley[17])

For Figure 16.48*a* and *b*, the depression angle and type are the parameters, whereas for Figure 16.48*c*, the terrain slope is the parameter. The lines are based on linear regression on the logarithmic frequency scale with the exception of desert and grass at very low depression angles where a quadratic regression was used. Note that urban, mountain, and forest tend to have relatively high values compared with agriculture and uncultivated land with moderate slopes.

16.10 IMAGING RADAR INTERPRETATION

Side-looking fine-resolution imaging radars with real or synthetic apertures produce images that closely resemble aerial photographs. Both shadows and differences in σ^0 for different parts of the ground produce image-intensity variations like those in photographs. For this reason, photointerpreters can easily learn to interpret radar images. However, since the radar images are due to microwave reflectivity, not optical reflectivity, the interpreters must understand the differences and that the images at the different wavelengths are, in fact, complementary. Moreover, the geometrical distortions for radar images are those of a side-looking range measurement system, whereas those of aerial photos are those of a down-looking angle measurement system—a difference that the interpreter must understand. At low grazing angles, the distortions are small for radar, but at low incidence angles, they are large. Moreover, the speckle in radar images is not present in photographs.

Modern imaging radars use digital recording and processing, and the images are produced on film or manipulated digitally. Because the side-looking configuration produces a strip image, the output films are usually in the form of long strips. Most cameras produce separate images that are approximately square. Strip film cameras and optical-infrared scanners produce strip images like those of radars, but with different distortions because they are angle-, not range-, measuring devices.

Every pure and applied science that uses aerial photography can also use radar images. This is particularly useful in cloudy environments, but radar is also useful even in clear weather because performance is independent of time of day so sun angle does not matter. Moreover, radar ground signatures differ from those in the visible and infrared. Radar has been applied to agriculture, forestry, geology, hydrology, urban geography, regional studies, oceanography, and ice mapping.

Fading complicates the interpretation of images by producing *speckle*. This means that averaging of the speckled image usually is necessary. Sometimes the processor does the averaging, and sometimes the interpreter does it mentally; it must be done to interpret an image. The image intensity for individual pixels with a single-look SAR follows the Rayleigh distribution (actually the exponential distribution if square-law detection is used). Most SAR processors sacrifice some spatial resolution by averaging, say, four pixels together after detection. Transmitting more bandwidth than needed for range resolution accomplishes this purpose without loss of needed spatial resolution,[180] but it takes more transmitter power. Suitable frequency agility can accomplish the same result.

Tradeoffs exist between spatial resolution and measurement precision. The latter can be used to define a *gray-level resolution*.[181] One can then think of image resolution in terms of a volume:

$$V = r_a r_y r_g \tag{16.33}$$

where r_a is the along-track resolution, r_y is the ground-range resolution, and r_g is the gray-level resolution. The referenced study showed that interpretability depends on V; so tradeoffs between the three elements of V are possible. Best results for a human interpreter occur when three independent samples of the fading are averaged. Ignoring this fading (speckle) can lead to erroneous conclusions on the spatial resolution needed for a given application.

Single-frequency, single-polarization radar images are useful. However, use of multiple polarizations (particularly including cross polarization) and multiple frequencies clearly increases their value. Different angles of incidence are most suitable for different applications. For example, soil moisture monitoring is best within 20° of vertical at frequencies near 5 GHz. Vegetation discrimination is better, however, at higher frequencies and angles of incidence. For some applications, use of the full polarization matrix, including the phase, is useful. The phase information is particularly helpful in assessing scattering mechanisms, as in forests.[32]

Because the literature in this field is so massive, the radar engineer wishing to learn more about the subject should consult the *Manual of Remote Sensing,*[40,41] *Microwave Remote Sensing*,[37] especially Vol. III and Chapter 11 of Vol. II, and the journals outlined early in Section 16.7.

REFERENCES

1. H. Goldstein, "Sea Echo," in *Propagation of Short Radio Waves*, D. E. Kerr (ed.), MIT Radiation Laboratory Series, Chap. 6, Vol. 13, New York: McGraw-Hill Book Company, 1951.
2. R. L. Cosgriff, W. H. Peake, and R. C. Taylor, "Terrain scattering properties for sensor system design," *Terrain Handbook II*, Columbus: The Ohio State University, Eng. Exp. Sta. Antenna Lab., 1959.
3. R. K. Moore, "Radar scattering cross-section per unit area and radar astronomy," *IEEE Spectrum*, p. 156, April 1966.
4. A. K. Fung, *Microwave Scattering and Emission Models and Their Applications*. Boston: Artech House, 1994.
5. A. G. Voronovich, *Wave Scattering from Rough Surfaces*. New York: Springer-Verlag, 1994.
6. G. Ruck, D. Barrick, W. Stuart, and C. Krichbaum, *Radar Cross Section Handbook*, New York: Plenum Press, 1968.
7. R. K. Moore, "Resolution of vertical incidence radar return into random and specular components," University of New Mexico, Eng. Exp. Sta., Albuquerque, 1957.
8. J. M. Banhart (ed.) *Remote Sensing Laboratory Publication List 1964–1980*, Lawrence: University of Kansas, Remote Sensing Lab., 1981.
9. J. M. Banhart (ed.), *Remote Sensing Laboratory Publication List 1981–1983*, Vol. TR-103, Lawrence: University of Kansas, Remote Sensing Lab., 1984.
10. G. P. de Loor, P. Hoogeboom, and E. P. W. Attema, "The Dutch ROVE program," *IEEE Trans.*, vol. GE-20, pp. 3–11, 1982.
11. B. A. M. Bouman and H. W. J. vanKasteren, *Ground-based X-band Radar Backscatter Measurements of Wheat, Barley and Oats*, Wageningen Netherlands: Center for Agrobiological Research, 1989.
12. T. LeToan, "Active microwave signatures of soil and crops: significant results of three years of experiments," *Dig. Int. Geosci. Remote Sensing Symp. (IGARSS '82), IEEE* 82CH14723-6, vol. 1, 1982.
13. Martinez, et al., "Measurements and Modeling of Vertical Backscatter Distribution in Forest Canopy," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 38, pp. 710–719, 2000.
14. R. Bernard and D. Vidal-Madjar, "C-band radar cross-section of the Guyana Rain Forest: possible use as a reference target for spaceborne radars," *Remote Sensing of Envir.*, vol. 27, pp. 25–36, 1989.

15. A. R. Edison, R. K. Moore, and B. D. Warner, "Radar return measured at near-vertical incidence," *IEEE Trans. Ant. & Prop.*, vol. AP-8, pp. 246–254, 1960.
16. S. P. Gogineni and K. Jezek, "Ultra-wideband radar measurements over bare and snow-covered saline ice," *Proc. IGARSS95*, vol. 2, pp. 859–861, 1995.
17. J. B. Billingsley, *Low-Angle Radar Land Clutter: Measurements and Empirical Models*. Norwich, NY: William Andrew Publishing, 2002.
18. R. E. Clapp, "A theoretical and experimental study of radar ground return," MIT Radiat. Lab. Rept. 6024, Cambridge, MA, 1946.
19. T. S. George, "Fluctuations of ground clutter return in airborne radar equipment," *Proc. IEE (London)*, vol. 99, pp. 92–99, 1952.
20. E. A. Reitz et al., "Radar terrain return study, final report: Measurements of terrain back-scattering coefficients with an airborne X-band radar," Goodyear Aerospace Corporation, *GERA-463*, Phoenix, 1959.
21. J. P. Campbell, "Back-scattering characteristics of land and sea at X band," in *Proc. Natl. Conf. Aeronaut. Electron.*, 1958.
22. F. C. MacDonald, "The correlation of radar sea clutter on vertical and horizontal polarization with wave height and slope," in *IRE. Conv. Rec.*, vol. 4, 1956, pp. 29–32.
23. W. S. Ament, F. C. MacDonald, and R. Shewbridge, "Radar terrain reflections for several polarizations and frequencies," in *Proc. Symp. Radar Return, NOTS TP2359*, U.S. Naval Ordnance Test Station, Test Station, China Lake, CA, 1959.
24. C. R. Grant and B. S. Yaplee, "Backscattering from water and land at centimeter and millimeter wavelengths," *Proc. IRE.*, vol. 45, pp. 972–982, 1957.
25. Guinard et al., "Variation of the NRCS of the sea with increasing roughness," *J. Geophys. Res.*, vol. 76, pp. 1525–1538, 1971.
26. C. E. K. Livingstone, P. Singh, and A. L. Gray, "Seasonal and regional variations of active/passive microwave signatures of sea ice," *IEEE Trans.*, vol. GE-25, pp. 159–173, 1987.
27. H. McNairn et al., "Identification of agricultural tillage practices from C-band radar backscatter," *Canadian Journal of Remote Sensing*, vol. 22, 1996, pp. 154–162.
28. R. W. Larson, R. E. Hamilton, and F. L. Smith, "Calibration of synthetic aperture radar, *Dig. IGARSS '81*, pp. 938–943, 1981.
29. C. E. K. Livingstone et al., "Springtime C-band SAR backscatter signatures of Labrador Sea marginal ice: measurements versus modeling predictions," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 29, pp. 29–41, 1991.
30. A. Haskell and B. M. Sorensen, "The European SAR-580 project," *Dig. IGARSS '82, IEEE 82CH*14723-6, Sess. WA-5, pp. 1.1–1.5, 1982.
31. D. N. Held, "The NASA/JPL multipolarization SAR aircraft program," *Dig. IGARSS 85*, pp. 454–457, 1985.
32. D. L. Evans et al., "Radar polarimetry: analysis tools and applications," *IEEE Trans. Geosc. & Rem. Sens.*, vol. 26, pp. 774–789, 1988.
33. Hoogeboom et al., "The PHARUS Project, Results of the Definition Study including the SAR Testbed PHARS," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 30, pp. 723–735, 1992.
34. Y.-L. Desnos et al., "The ENVISAT advanced synthetic aperture radar system," *Proc. IGARSS2000*, vol. 3, pp. 1171–1173, 2000.
35. P. Fox, A. P. Luscombe, and A. A. Thompson, "RADARSAT-2 SAR modes development and utilization," *Canadian Jour. of Rem. Sens.*, vol. 30, pp. 258–264, 2004.
36. H. Wakabayashi et al., "Airborne L-band SAR system: Characteristics and initial calibration results," *Proc IGARSS'99*, vol. 1, pp. 464–466, 1999.
37. F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing: Active and Passive*, Vol. I and Vol. II, Reading, MA: Addison-Wesley Publishing Company, 1981 and 1982; Vol. III, Norwood, MA: Artech House, 1986.
38. F. T. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, Norwood, MA: Artech House, 1989.

39. M. W. Long, *Radar Reflectivity of Land and Sea*, 2nd Ed., Norwood, MA: Artech House, 1983.
40. R. N. Colwell, D. S. Simonett, J. E. Estes, F. T. Ulaby, G. A. Thorley, et al., *Manual of Remote Sensing*, 2nd Ed., Vols. I and II, Falls Church, VA: American Society of Photogrammetry, 1983.
41. F. M. Henderson and A. J. Lewis, *Manual of Remote Sensing, Principles and Applications of Imaging Radar*, Vol. 2, 3rd Ed., New York: John Wiley & Sons, 1998.
42. J. R. Lundien, "Terrain analysis by electromagnetic means: radar responses to laboratory prepared soil samples," *U.S. Army Waterways Exp. Sta., TR 3-639*, Vicksburg, MS, 1966.
43. L. K. Wu, R. K. Moore, R. Zoughi, F. T. Ulaby, and A. Afifi, "Preliminary results on the determination of the sources of scattering from vegetation canopies at 10 GHz," pts. I and II, *Int. J. Remote Sensing*, vol. 6, pp. 299–313, 1985.
44. L. K. Wu, R. K. Moore, and R. Zoughi, "Sources of scattering from vegetation canopies at 10 GHz," *IEEE Trans.*, vol. GE-23, pp. 737–745, 1985.
45. R. Zoughi, J. Bredow, and R. K. Moore, "Evaluation and comparison of dominant backscattering sources at 10 GHz in two treatments of tall-grass prairie," *Remote Sensing Environ.*, vol. 22, pp. 395–412, 1987.
46. R. Zoughi, L. K. Wu, and R. K. Moore, "Identification of major backscattering sources in trees and shrubs at 10 GHz," *Remote Sensing Environ.*, vol. 19, pp. 269–290, 1986.
47. J. F. Paris, "Probing thick vegetation canopies with a field microwave spectrometer," *IEEE Trans.*, vol. GE-24, pp. 886–893, 1986.
48. S. T. Wu, "Preliminary report on measurements of forest canopies with C-Band radar scatterometer at NASA/NSTL," *IEEE Trans.*, vol. GE-24, November 1986.
49. D. E. Pitts, G. D. Badhwar, and E. Reyna, "The Use of a helicopter mounted ranging scatterometer for estimation of extinction and scattering properties of forest canopies," *IEEE Trans.*, vol. GE-26, pp. 144–152, 1988.
50. R. Bernard, M. E. Frezal, D. Vidal-Madjar, D. Guyon, and J. Riom, "Nadir looking airborne radar and possible applications to forestry," *Remote Sensing Environ.*, vol. 21, pp. 297–310, 1987.
51. S. L. Durden, J. D. Klein, and H. A. Zebker, "Polarimetric radar measurements of a forested area near Mt. Shasta," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 29, pp. 444–450, 1991.
52. D. K. Barton, "Land clutter models for radar design and analysis," *Proc. IEEE*, vol. 73, pp. 198–204, 1985.
53. R. K. Moore, *Traveling Wave Engineering*, New York: McGraw-Hill Book Company, 1960.
54. A. H. Schooley, "Upwind-downwind ratio of radar return calculated from facet size statistics of wind disturbed water surface," *Proc. IRE*, vol. 50, pp. 456–461, 1962.
55. D. O. Muhleman, "Radar scattering from venus and the moon," *Astron. J.*, vol. 69, pp. 34–41, 1964.
56. A. K. Fung, "Theory of cross polarized power returned from a random surface," *Appl. Sci. Res.*, vol. 18, pp. 50–60, 1967.
57. I. Katz and L. M. Spetner, "Two statistical models for radar return," *IRE Trans.*, vol. AP-8, pp. 242–246, 1960.
58. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, New York: Macmillan Company, 1963.
59. P. Beckmann, "Scattering by composite rough surfaces," *Proc. IEEE*, vol. 53, pp. 1012–1015, 1965.
60. A. K. Fung and H. J. Eom, "An approximate model for backscattering and emission from land and sea," *Dig. IGARSS '81*, vol. I, pp. 620–628, 1981.
61. H. S. Hayre and R. K. Moore, "Theoretical scattering coefficients for near-vertical incidence from contour maps," *J. Res. Nat. Bur. Stand.*, vol. 65D, pp. 427–432, 1961.
62. H. Davies, "The reflection of electromagnetic waves from a rough surface," *Proc. IEE (London)*, pt. 4, vol. 101, pp. 209–214, 1954.
63. A. K. Fung and R. K. Moore, "The correlation function in Kirchoff's method of solution of scattering of waves from statistically rough surfaces," *J. Geophys. Res.*, vol. 71, pp. 2929–2943, 1966.

64. J. V. Evans and G. H. Pettengill, "The scattering behavior of the moon at wavelengths of 3.6, 68, and 784 centimeters," *J. Geophys. Res.*, vol. 68, pp. 423–447, 1963.
65. J. W. Wright, "A new model for sea clutter," *IEEE Trans*., vol. AP-16, pp. 217–223, 1968.
66. F. G. Bass, I. M. Fuks, A. I. Kalmykov, I. E. Ostrovsky, and A. D. Rosenberg, "Very high frequency radiowave scattering by a disturbed sea surface," *IEEE Trans.*, vol. AP-16, pp. 554–568, 1968.
67. S. O. Rice, "Reflection of electromagnetic waves by slightly rough surfaces," *Commun. Pure Appl. Math.*, vol. 4, pp. 351–378, 1951.
68. Ref. 37, vol. II, p. 961.
69. Ref. 37, vol. II, chap. 12.
70. Ref. 37, vol. III, chap. 13.
71. R. H. Lang and J. S. Sidhu, "Electromagnetic scattering from a layer of vegetation: a discrete approach," *IEEE Trans.*, vol. GE-21, pp. 62–71, 1983.
72. A. K. Fung, "A review of volume scatter theories for modeling applications," *Radio Sci.*, vol. 17, pp. 1007–1017, 1982.
73. Y. S. Kim, "Theoretical and experimental study of radar backscatter from sea ice," Ph.D. dissertation, University of Kansas, Lawrence, 1984.
74. J. M. Stiles and K. Sarabandi, "Electromagnetic scattering from grassland—Part I: A fully phase-coherent scattering model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 339–348, 2000.
75. H. O. Rydstrom, "Interpreting local geology from radar imagery," *Bull. Geol. Soc. Am.*, vol. 78, pp. 429–436, 1967.
76. W. K. Lee, "Analytical investigation of urban SAR features having a group of corner reflectors," *IGARSS 2001*, vol. 3, pp. 1262–1264, 2001.
77. M. F. Chen and A. K. Fung, "A study of the validity of the integral equation model by moment method simulation—cylindrical case," *Remote Sensing of Envir.*, vol. 29, pp. 217–228, 1989.
78. A. K. Fung, M. R. Shah, and S. Tjuatja, "Numerical simulation of scattering from three-dimensional randomly rough surfaces," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 32, pp. 986–994, 1994.
79. S. O. Rice, "Mathematical analysis of random noise," pt. 1, *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, 1944; pt. II, vol. 24, pp. 46–156, 1945.
80. G. A. Shmidman, "Generalized radar clutter model," *IEEE Trans. on Aerosp. Elec. Sys.*, vol. 35, pp. 857–865, 1999.
81. R. D. DeRoo et al., "MMW scattering characteristics of terrain at near-grazing incidence," *IEEE Trans. on Aerosp. Elec. Sys.*, vol. 35, pp. 1010–1018, 1999.
82. J. B. Billingsley et al., "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. & Electron. Sys.* vol. 35, pp. 579–593, 1999.
83. Ref. 37, vol. II, pp. 487–492.
84. F. T. Ulaby, W. H. Stiles, D. Brunfeldt, and E. Wilson, "1-35 GHz microwave scatterometer," in *Proc. IEEE/MTT-S, Int. Microwave Symp.*, IEEE 79*CH*1439-9 *MIT-S*, 1979.
85. D. R. Brunfeldt and F. T. Ulaby, "An active radar calibration target," *Dig. IGARSS '82*, IEEE 82*CH*14723-6, 1982.
86. A. Freeman, Y. Shen, and C. L. Werner, "Polarimetric SAR calibration experiment using active radar calibrators," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 28, pp. 224–240, 1990.
87. Ref. 37, vol. II, pp. 766–779.
88. F. J. Janza, "The analysis of a pulse radar acquisition system and a comparison of analytical models for describing land and water radar return phenomena," Ph.D. dissertation, University of New Mexico, Albuquerque, 1963.
89. F. J. Janza, R. K. Moore, and R. E. West, "Accurate radar attenuation measurements achieved by inflight calibration," *IEEE Trans.*, vol. PGI-4, pp. 23–30, 1955.
90. E. M. Bracalente, W. L. Jones, and J. W. Johnson, "The Seasat—a satellite scatterometer," *IEEE Trans.*, vol. OE-2, pp. 200–206, 1977.

91. F. K. Li, D. Callahan, D. Lame, and C. Winn, "NASA scatterometer on NROSS—a system for global observations on ocean winds," *Dig. IGARSS '84*, 1984.
92. R. K. Moore and W. J. Pierson, "Measuring sea state and estimating surface winds from a polar orbiting satellite," in *Proc. Int. Symp. Electromagn. Sensing of Earth from Satellites*, 1965, pp. R1–R26.
93. L. J. Cote et al., "The directional spectrum of a wind-generated sea as determined from data obtained by the stereo wave observation project," *New York University Meterorol. Pap.*, vol. 2, no. 66, 1960.
94. S. P. Gogineni et al., "Application of plane waves for accurate measurement of microwave scattering from geophysical surfaces," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 627–633, 1995.
95. T. F. Bush and F. T. Ulaby, "8–18 GHz radar spectrometer," University of Kansas, Remote Sensing Lab., vol. TR 177-43, Lawrence, September 1973.
96. Ref. 37, vol. II, pp. 779–791; vol. III, chap. 14.
97. R. Zoughi, L. K. Wu, and R. K. Moore, "SOURCESCAT: A very fine resolution radar scatterometer," *Microwave J.*, vol. 28, pp. 183–196, 1985.
98. S. P. Gogineni, F. A. Hoover, and J. W. Bredow, "High-performance, inexpensive polarimetric radar for in situ measurements," *Proc. IGARSS89*, vol. 28, pp. 450–455, 1990.
99. R. K. Moore, "Effect of pointing errors and range on performance of dual-pencil-beam scatterometers," *IEEE Trans.*, vol. GE-23, pp. 901–905, 1985.
100. A. R. Edison, "An acoustic simulator for modeling backscatter of electromagnetic waves," Ph.D. dissertation, University of New Mexico, Albuquerque, 1961.
101. B. E. Parkins and R. K. Moore, "Omnidirectional scattering of acoustic waves from rough surfaces of known statistics," *J. Acoust. Soc. Am.*, vol. 50, pp. 170–175, 1966.
102. R. K. Moore, "Acoustic Simulation of radar returns," *Microwaves*, vol. 1, no. 7, pp. 20–25, 1962.
103. M. C. Dobson, F. T. Ulaby, D. R. Brunfeldt, and D. N. Held, "External calibration of SIR-B imagery with area-extended and point targets," *IEEE Trans.*, vol. GE-24, pp. 453–461, 1986.
104. D. Vaillant and A. Wadsworth, "Preliminary results of some remote sensing campaigns of the French Airborne SAR VARAN-S," *Dig. IGARSS '86*, pp. 495–500, 1986.
105. H. Hirosawa and Y. Matsuzaka, "Calibration of cross-polarized SAR imagery using dihedral corner reflectors," *Dig. IGARSS '86*, pp. 487–492, 1986.
106. D. R. Brunfeldt and F. T. Ulaby, "Active reflector for radar calibration," *IEEE Trans.*, vol. GE-22, pp. 165–169, 1984.
107. P. Hartl, M. Reich, and S. Bhagavathula, "An attempt to calibrate air-borne SAR image using active radar calibrators and ground-based scatterometers," *Dig. IGARSS 86*, pp. 501–508, 1986.
108. R. W. Larson et al., "Bistatic clutter measurements," *IEEE Trans.*, vol. AP-26, pp. 801–804, 1978.
109. J. Renau and J. A. Collinson, "Measurements of electromagnetic backscattering from known rough surfaces," *Bell Syst. Tech. J.*, vol. 44, pp. 2203–2226, 1965.
110. D. Kieu, "Effect of tall structures on microwave communication systems," M.S. thesis, University of Kansas, Lawrence, 1988.
111. F. T. Ulaby et al., "Millimeter-wave bistatic scattering from ground and vegetation targets," *IEEE Trans. Geosc. & Rem. Sens.*, vol. GE-26, pp. 229–243, 1988.
112. T.-K. Chan et al., "Experimental studies of bistatic scattering from two-dimensional conducting random rough surfaces," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 34, pp. 674–680, 1996.
113. W. H. Stiles, D. Brunfeldt, and F. T. Ulaby, "Performance analysis of the MAS (Microwave Active Spectrometer) systems: calibration, precision and accuracy," University of Kansas, Remote Sensing Lab., vol. TR 360-4, Lawrence, 1979.
114. F. T. Ulaby et al., "1-35 GHz microwave scatterometer," *Proc. IEEE/MTT-S 1979 Intl. Microwave Symp.*, vol. '79CH1439-9 MTT-S', 1979.

115. R. K. Moore, K. A. Soofi, and S. M. Purduski, "A radar clutter model: average scattering coefficients of land, snow, and ice," *IEEE Trans.*, vol. AES-16, pp. 783–799, 1980.

116. R. K. Moore et al., "Simultaneous active and passive microwave response of the Earth—the Skylab RADSCAT experiment," in *Proc. Ninth Int. Symp. Remote Sensing Environ.*, University of Michigan, Ann Arbor, 1974, pp. 189–217.

117. Ref. 21. See summaries in vol. II, chap. 11, and vol. III, chap. 21.

118. F. T. Ulaby, "Vegetation clutter model," *IEEE Trans.*, vol. AP-28, pp. 538–545, 1980.

119. W. H. Stiles and F. T. Ulaby, "The active and passive microwave response to snow parameters, part I: wetness," *J. Geophys. Res.*, vol. 85, pp. 1037–1044, 1980.

120. F. T. Ulaby and W. H. Stiles, "The active and passive microwave response to snow parameters, part II: water equivalent of dry snow," *J. Geophys. Res.*, vol. 85, pp. 1045–1049, 1980.

121. I. J. Birrer, E. M. Bracalante, G. J. Dome, J. Sweet, and G. Berthold, "Signature of the Amazon rain forest obtained with the Seasat scatterometer," *IEEE Trans.*, vol. GE-20, pp. 11–17, 1982.

122. R. Bernard and D. Vidal-Madjar, "C-band radar cross-section of the Guyana rain forest: possible use as a reference target for spaceborne radars," *Remote Sensing of Envir.*, vol. 27, pp. 25–36, 1989.

123. R. K. Moore and M. Hemmat, "Determination of the vertical pattern of the SIR-B antenna," *Int'l Jour. Rem. Sens.*, vol. 9, pp. 839–847, 1988.

124. M. Shimada, "Long-term stability of L-band normalized radar cross section of Amazon rainforest using the JERS-1 SAR," *Canadian Jour. of Rem. Sens.*, vol. 31, pp. 132–137, 2005.

125. A. R. Edison, R. K. Moore, and B. D. Warner, "Radar return measured at near-vertical incidence," *IEEE Trans.*, vol. AP-8, pp. 246–254, 1960.

126. C. H. Bidwell, D. M. Gragg, and C. S. Williams: "Radar return from the vertical for ground and water surface," Sandia Corporation, Albuquerque, NM, 1960.

127. Y. S. Kim, R. K. Moore, R. G. Onstott, and S. P. Gogineni, "Towards identification of optimum radar parameters for sea-ice monitoring," *J. Glaciol.*, vol. 31, pp. 214–219, 1985.

128. T. LeToan et al., "Multitemporal and dual-polarization observations of agricultural vegetation covers by X-band SAR images," *IEEE Trans. on Geosc. and Remote Sensing*, vol. GE-27, pp. 709–718, 1989.

129. B. A. M. Bouman and H. W. J. vanKasteren, *Ground-based X-band Radar Backscatter Measurements of Wheat, Barley and Oats*, Wageningen NETHERLANDS: Center for Agrobiological Research, 1989.

130. B. Brisco, R. J. Brown, and G. J. Sofko, "The CCRS ground-based microwave facility," *IGARSS88*, vol. 1, pp. 575–576, 1988.

131. E. Stotzer, V. Wegmuller, R. Huppi, and C. Matzler, "Dielectric and surface parameters related to microwave scatter and emission properties," *Dig. IGARSS '86*, pp. 599–609, 1986.

132. T. Nagler and H. Rott, "Retrieval of wet snow by means of multitemporal SAR data," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 38, pp. 754–765, 2000.

133. F. J. Janza, R. K. Moore, and B. D. Warner, "Radar cross-sections of terrain near vertical incidence at 415 Mc, 3800 Mc, and extension of analysis to X band," University of New Mexico, Eng. Exp. Sta., TR EE-21, Albuquerque, 1959.

134. Ref. 37, vol. III, Fig. 21.20, p. 1825.

135. Ref. 37, vol. III, Fig. 21.22, p. 1827.

136. Ref. 37, vol. III, Fig. 21.41, p. 1856.

137. D. H. Hoekman, "Radar backscattering of forest stands," *Int. J. Remote Sensing*, vol. 6, pp. 325–343, 1985.

138. D. H. Hoekman et al., "Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Columbian Amazon," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 38, pp. 685–696, 2000.

139. M. C. Dobson and F. T. Ulaby, "Microwave backscatter dependence on surface roughness, soil moisture and soil texture: Part III—soil tension," *IEEE Trans.*, vol. GE-19, pp. 51–61, 1981.

140. T. J. Schmugge, "Effect of texture on microwave emission from soils," *IEEE Trans.*, vol. GE-18, pp. 353–361, 1980.

141. F. T. Ulaby, A. Aslam, and M. C. Dobson, "Effects of vegetation cover on the radar sensitivity to soil moisture," University of Kansas, Remote Sensing Lab., TR 460-10, Lawrence, 1981.

142. M. C. Dobson, F. Kouyate, and F. T. Ulaby, "A reexamination of soil textural effects on microwave emission and backscattering," *IEEE Trans.*, vol. GE-22, pp. 530–535, 1984.

143. F. T. Ulaby, B. Brisco, and M. C. Dobson, "Improved spatial mapping of rainfall events with spaceborne SAR imagery," *IEEE Trans.*, vol. GE-21, pp. 118–121, 1983.

144. F. T. Ulaby, M. C. Dobson, J. Stiles, R. K. Moore, and J. C. Holtzman, "A simulation study of soil moisture estimation by a space SAR," *Photogramm. Eng. Remote Sensing*, vol. 48, pp. 645–660, 1982.

145. Z. Li et al., "Soil moisture measurement and retrieval using envisat asar imagery," *Proc. IGARSS04*, vol. V, pp. 3539–3542, 2004.

146. J. Shi et al., "Estimation of soil moisture with L-band multipolarization radar," *Proc. IGARSS04*, vol. II, pp. 815–818, 2004.

147. Y. Kim and J. van Zyl, "Vegetation effects on soil moisture estimation," *Proc. IGARSS04*, vol. II, pp. 800–802, 2004.

148. E. Attema and F. T. Ulaby, "Vegetation modeled as a water cloud," *Radio Sci.*, vol. 13, pp. 357–364, 1978.

149. Ref. 37, vol. III, p. 1873.

150. H. Eom and A. K. Fung, "A scatter model for vegetation up to K_u-band," *Remote Sensing Environ.*, vol. 15, pp. 185–200, 1984.

151. W. H. Stiles and F. T. Ulaby "The active and passive microwave response to snow parameters, Part I: Wetness," *J. Geophys. Res.*, vol. 85, pp. 1037–1044, 1980.

152. W. H. Stiles, F. T. Ulaby, A. K. Fung, and A. Aslam, "Radar spectral observations of snow," *Dig. IGARSS '81*, pp. 654–668, 1981.

153. A. V. Bushuyev, N. A. Volkov, and V. S. Loshchilov, *Atlas of Ice Formations*, Leningrad: Gidrometeoizdat, 1974. (In Russian with English annotations.)

154. A. L. Gray, R. K. Hawkins, C. E. Livingstone, L. D. Arsenault, and W. M. Johnstone, "Simultaneous scatterometer and radiometer measurements of sea ice microwave signatures," *IEEE J.*, vol. OE-7, pp. 20–32, 1982.

155. V. S. Loshchilov and V. A. Voyevodin, "Determining elements of drift of the ice cover and movement of the ice edge by the aid of the 'Toros' side scanning radar station," *Probl. Arktiki Antarkt* (in Russian), vol. 40, pp. 23–30, 1972.

156. S. Haykin et al., *Remote Sensing of Sea Ice and Icebergs*, New York: Wiley-IEEE, 1994.

157. R. K. Raney et al., "RADARSAT," Proc. IEEE, vol. 79, pp. 839–849, 1991.

158. B. Ramsay et al., " Use of RADARSAT data in the Canadian ice service," *Canadian Journal of Remote Sensing*, vol. 24, pp. 36–42, 1998.

159. G. I. Belchansky and D. C. Douglas, "Seasonal comparisons of sea ice concentration estimates derived from SSM/I, OKEAN, and RADARSAT data," *Rem. Sens. Envir.*, vol. 81, pp. 67–81, 2002.

160. M. Nazirov, A. P. Pichugin, and Y. G. Spiridonov, *Radiolokatsia Poverchnosti Zemli Iz-Kosmoca (Radar Observation of the Earth from Space)*, Leningrad: Hydrometeoizdat, 1990. (In Russian.)

161. Mitnik et al., "Structure and dynamics of the Sea of Okhotsk marginal ice zone from 'ocean' satellite radar sensing data," *J. Geophys. Res.*, vol. 97, pp. 7249–7445, 1992.

162. M. R. Drinkwater, R. Hosseinmostafa, and S. P. Gogineni, "C-band backscatter measurements of winter sea-ice in the Weddell Sea, Antarctica," *International Journal of Remote Sensing*, vol. 16, pp. 3365, 1995.

163. F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications*. Boston: Artech House, 1990.

164. M. A. Sletten and D. J. McLaughlin, "Radar polarimetry," in *Wiley Encyclopedia of Electrical and Electronics Engineering Online*, J. Webster (ed.), New York: John Wiley & Sons, Inc., 1999.

165. J. van Zyl and Y. Kim, "Remote sensing by radar," in *Wiley Encyclopedia of Electrical and Electronics Engineering Online*, J. Webster (ed.), New York: John Wiley & Sons, Inc., 1999.

166. W. M. Boerner et al., "On the basic principles of radar polarimetry: the target characteristic polarization state theory of Kennaugh, Huynen's polarization fork concept, and its extension to the partially polarized case," *Proc. IEEE*, vol. 79, pp. 1538–1550, 1990.

167. J. J. van Zyl, H. Zebker, and D. N. Held, "Imaging radar polarization signatures: Theory and observation," *Radio Sci.*, vol. 22, pp. 529–543, 1987.

168. S. A. Morain and D. S. Simonett, "K-band radar in vegetation mapping," *Photog. Eng. and Rem. Sens.*, vol. 33, pp. 730–740, 1967.

169. P. C. Dubois and J. van Zyl, "Polarization filtering of SAR data," *Digest IGARSS88*, vol. 3, pp. 1816–1819, 1989.

170. A. A. Swartz et al., "Optimal polarization for achieving maximum contrast in radar images," *J. Geophys. Res.*, vol. 93, pp. 15252–15260, 1988.

171. S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 35, pp. 68–78, 1997.

172. J. van Zyl, "Unsupervised classification of scattering behavior using radar polarimetry data," *IEEE Trans. Geosc. Rem. Sens.*, vol. 27, pp. 36–45, 1989.

173 Touzi et al., "Polarimetric discriminators for SAR images," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 30, pp. 973–980, 1992

174. S. L. Durden, J. D. Klein, and H. A. Zebker, "Polarimetric radar measurements of a forested area near Mt. Shasta," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 29, pp. 111–450, 1991.

175 Hoekman et al., "Biophysical forest type characterization in the Columbian Amazon by airborne polarimetric SAR," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 40, pp. 1288–1300, 2002.

176. P. Ferrazzoli et al., "The potential of multifrequency polarimetric SAR in assessing agricultural and arboreous biomass," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 35, pp. 5–17, 1997.

177. Inoue et al., "Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables," *Remote Sensing of Envir.*, vol. 81, pp. 194–204, 2002.

178. S. V. Nghiem et al., "Polarimetric signatures of sea ice, 2, experimental observations," *J. Geophys. Res.*, vol. 100, pp. 13681–13698, 1995.

179. A. Martini, L. Ferro-Famil, and E. Pottier, "Multi-frequency polarimetric snow discrimination in alpine areas," *Proc. IGARSS04*, vol. VI, pp. 3684–3687, 2004.

180. L. L. Hess et al., "Delineation of inundated area and vegetation along the Amazon floodplain with the SIR-C synthetic aperture radar," *IEEE Trans. on Geosc. and Remote Sensing*, vol. 33, pp. 896–904, 1995.

181. R. K. Moore, W. P. Waite, and J. W. Rouse, "Panchromatic and polypanchromatic radar," *Proc. IEEE*, vol. 57, pp. 590–593, 1969.

182. R. K. Moore, "Tradeoff between picture element dimensions and noncoherent averaging in side-looking airborne radar," *IEEE Trans.*, vol. AES-15, pp. 696–708, 1979.

Chapter 17
Synthetic Aperture Radar*

Roger Sullivan

Institute for Defense Analyses

Most of the discussion in this Handbook concerns *real aperture radar (RAR)*, where the antenna is a physical object that first emits, and then collects, the radiation. We turn our attention to the case where the antenna moves to cover a *synthetic aperture*, thus producing *synthetic aperture radar (SAR)*. This overview is based on Sullivan[1] and Cutrona[2†]; more detailed treatments are also provided in the literature.[3–11]

17.1 BASIC PRINCIPLE OF SAR

For airborne or spaceborne ground-mapping radar, there has been continuous desire to achieve finer resolution. We shall use the term *range resolution* to mean the resolution along the line-of-sight (LOS) from the radar to the target region and *crossrange resolution* to mean the resolution along the direction perpendicular to the LOS and parallel to the ground. The former is also frequently termed *downrange resolution* to emphasize that it is along the LOS. Crossrange resolution is also frequently called *azimuth resolution*, since it is measured along a line obtained by holding range constant and varying the azimuth (as measured from the physical antenna) of the LOS. If (and only if) the LOS remains perpendicular to the direction of flight, range resolution is sometimes called *cross-track resolution*, and crossrange resolution is sometimes called *along-track resolution*.

With respect to SAR resolution, the preferred terms are *fine* and *coarse*. Better resolution is *finer*, not *greater*; poorer resolution is *coarser*, not *less*. In this way ambiguity in terminology can be avoided. Of course, in practice the terms *high-resolution* (fine resolution) and *low-resolution* (coarse resolution) are often used without ambiguity.

Crossrange resolution was initially achieved by use of a narrow beam. The beamwidth (in radians) θ_B of an aperture antenna is given approximately by the wavelength

* The present chapter draws significantly from Dr. Sullivan's book, *Radar Foundations for Imaging and Advanced Concepts*, Raleigh, NC: SciTech Publishing, 2004. Dr. Sullivan is grateful to SciTech for permission to quote considerable material from the chapters on SAR. He is also grateful to Mr. Michael Tuley (Institute for Defense Analyses) for reviewing the manuscript prior to publication and suggesting many improvements.

† In the first two editions of this Handbook, the chapter on SAR was written by Dr. Louis J. Cutrona, who is now deceased. In 1988, Dr. Sullivan had the privilege of working with Dr. Cutrona at the Environmental Research Institute of Michigan, Ann Arbor (now General Dynamics, Ypsilanti), and is grateful for having learned much from him concerning SAR.

λ divided by the aperture diameter: $\theta_B \approx \lambda/D$. The corresponding linear crossrange resolution at range R is then

$$\delta_{cr} \approx \frac{R\lambda}{D} \qquad \text{(real aperture)} \tag{17.1}$$

For example, if $\lambda = 3$ cm (X band) and $D = 2$ m, $\theta_B \approx 0.015$ radians. At a range of $R =$ 100 km, the crossrange resolution would be $R\,\theta_B \approx 1.5$ km, hardly fine enough to resolve such targets as buildings and vehicles. However, by using appropriate coherent processing, an aperture of modest size may be moved (e.g., by an aircraft or spacecraft—referred to as the *platform*) along a path in space—a *synthetic aperture*—and may achieve crossrange resolution comparable to what would, in principle, be achieved by a real aperture with a length equal to the path length (synthetic aperture) L_{SA}:

$$\delta_{cr} \approx \frac{R\lambda}{2L_{SA}} \approx \frac{\lambda}{2\Delta\theta} \qquad \text{(synthetic aperture)} \tag{17.2}$$

where $\Delta\theta$ is the synthetic aperture angle, i.e., the angle subtended by the synthetic aperture as seen from the target area. The additional factor of two in the denominator (compared with Eq. 17.1) occurs because of the SAR processing and will be discussed. For example, for a path length of 5 km, the crossrange resolution for the case discussed above would be ≈ 30 cm, clearly a superb improvement over the real-aperture case.

17.2 EARLY HISTORY OF SAR

The original concept of SAR was first described by Carl Wiley of Goodyear in 1951.[12] He called it *doppler beam sharpening (DBS)*. Later, the doppler beam sharpening mode of SAR was introduced as the name for a variable squint angle mode that produced a partial plan-position indicator (PPI) display. Thus, DBS really has two meanings in SAR: (1) the name for SAR invented by Wiley before it was called SAR and (2) the name for a PPI-like mode based on squinted SAR.

In 1952, the University of Illinois demonstrated the SAR concept. In 1953, during a summer study that launched a program known as Project Michigan, ideas related to synthetic apertures were discussed by L. J. Cutrona of the University of Michigan Willow Run Laboratories, C. W. Sherwin of the University of Illinois, W. Hausz of General Electric, and J. Koehler of Philco Corporation.[2] This resulted in a successful SAR program by the Michigan group.[13] The Illinois group also demonstrated successful SAR imaging.[14] The work of Cutrona et al.[13] and Sherwin et al.,[14] plus many other early SAR papers, are compiled in a very useful book by Kovaly.[15] More detailed histories of SAR development are given by Curlander and McConough,[4] Jackson and Apel,[10] and Ausherman et al.[11]

17.3 TYPES OF SAR

When we refer to SAR, we usually mean *focused SAR;* the term indicates that the phase information has been optimally processed to produce resolution comparable to the theoretical limit. In the development of SAR, several interesting techniques

preceded the development of focused SAR. These techniques are discussed in order of progressively finer resolution.

Precursors to Focused SAR.

Side-Looking Airborne Radar (SLAR). SLAR consisted of an aircraft-mounted RAR pointed perpendicular to the direction of flight (hence "side-looking"), with crossrange resolution ~ $R\lambda/D$.

Doppler Beam Sharpening (DBS). As previously mentioned, when Wiley first conceived of the concept that we now call SAR, he termed it *doppler beam sharpening*. Wiley explained it as follows: "I had the luck to conceive of the basic idea, which I called Doppler Beam Sharpening (DBS), rather than Synthetic Aperture Radar (SAR). Like all signal processing, there is a dual theory. One is a frequency-domain explanation. This is Doppler Beam Sharpening. If one prefers, one can analyze the system in the time domain instead. This is SAR. The equipment remains the same–just the explanation changes. Conception was reported in a Goodyear Aircraft report in 1951."[12]

Subsequently, as described in Section 8.1 of Schleher,[16] DBS came to refer to an airborne scanning mode in which the echoes from the scanning real beam are doppler-processed to produce crossrange resolution finer than that provided by the real beam alone; broadside crossrange resolution is $\approx R\lambda/2L_{\mathrm{DBS}}$, where L_{DBS} is the synthetic aperture length generated during a target dwell. Stimson[5] explains: "Typically, the antenna continuously scans the region of interest...Because the integration time is limited to the length of time a ground patch is in the antenna beam—or, if you prefer, the length of the array that can be synthesized is so limited—the resolution is coarser than can be achieved with a non-scanning antenna" (p. 434).

Unfocused SAR. This type of early SAR is described by Cutrona[2] as follows: "The coherent signals received at the synthetic array points are integrated, with no attempt made to shift the phases of the signals before integration. This lack of phase adjustment imposes a maximum upon the synthetic antenna length that can be generated. This maximum synthetic antenna length occurs at a given range when the round-trip distance from a radar target to the center of the synthetic array differs by $\lambda/4$ from the round-trip distance between the radar target and the extremities of the array." Cutrona shows that the crossrange resolution is approximately $\frac{1}{2}(\lambda R)^{\frac{1}{2}}$.[2]

Unfocused SAR is typically not used today and is included here only for historical reasons. It was used in the early days of SAR since the technology of that time did not support focused SAR.

Types of Focused SAR. In focused SAR, a phase correction is made for each returning pulse echo. This essentially results in the theoretical crossrange resolution of Eq. 17.2.

Stripmap SAR. Stripmap SAR (or "strip" SAR) is also sometimes called "search" SAR, since it is useful for imaging large areas at relatively coarse resolution. In stripmap SAR, the beam remains normal to the flight path (the latter is assumed to be a straight line at constant altitude) and continuously observes a *swath* (or *strip*) of terrain parallel to the flight path extending from some minimum range R_{min} to some maximum range R_{max} from the flight path.

For stripmap SAR, the synthetic aperture angle $\Delta\theta$ is essentially equal to the real-aperture beamwidth θ_B:

$$\Delta\theta \approx \theta_B \approx \frac{\lambda}{D} \tag{17.3}$$

Thus

$$\delta_{\text{cr}} \approx \frac{\lambda}{2\Delta\theta} \approx \frac{D}{2} \tag{17.4}$$

Under ideal conditions, as long as $D \gg \lambda$ and signal-to-noise ratio (SNR) $\gg 1$, then the smaller the physical antenna, the finer the crossrange resolution, independent of range.

As the physical antenna moves along the synthetic aperture, the return from a point target at a particular range will exhibit a quadratic phase behavior (i.e., phase varies as the square of the time referenced to the closest approach) that is unique to the target's location on the ground.[2,17] Some stripmap SARs use a filtering approach to take advantage of this phenomenon.[11] In fact, for the echo from a point target in the scene, a close analogy exists between its quadratic phase variation during a single pulse (from a linear FM [LFM] pulse echo) and its quadratic phase variation over many pulses due to platform motion (Stimson,[5] p. 421). Other stripmap SARs divide the strip into *subpatches* and use spotlight-SAR processing (next section) for each subpatch[3] (see also Section 4.8 of Ausherman et al.[11]).

The newer *Range Migration Algorithm* (RMA) (see Chapter 10 of Carrara et al.[3]), originally developed for seismic applications, provides the most theoretically correct solution to the stripmap image problem. It does not make a far-field approximation but treats the wavefronts as spherical. It is particularly applicable to SAR with very wide fractional bandwidth and/or wide synthetic aperture angle. RMA involves substantial computational complexity; however, as processors become more sophisticated, this limitation is disappearing. A simpler, faster version of RMA is the *Chirp-Scaling Algorithm* (CSA) (see Chapter 11 of Carrara et al.[3]).

Squinted Stripmap SAR. In this case, the antenna boresight is not normal to the flight path. As seen from a top view, the *squint angle* θ_{sq} is the angle between the antenna boresight and the normal to the flight path; thus, for a broadside beam, $\theta_{\text{sq}} = 0$ and $\delta_{\text{cr}} \cong D/2$. More generally,

$$\begin{aligned} \Delta\theta \cong \theta_B &\cong \frac{\lambda}{D\cos\theta_{\text{sq}}} \\ \delta_{\text{cr}} &\cong \frac{\lambda}{2\Delta\theta} \cong \frac{D\cos\theta_{\text{sq}}}{2} \end{aligned} \tag{17.5}$$

It is assumed here that the synthetic aperture length $L_{\text{SA}} \ll R$ and θ_{sq} is essentially constant during a data collection. Depending on details, this condition is valid only for θ_{sq} less than about 45°, since, for a given crossrange resolution, as θ_{sq} increases, L_{SA} increases, and the condition becomes no longer valid.

Spotlight SAR. Spotlight SAR (sometimes called "spot SAR") is used to obtain a relatively fine-resolution image of a known location or target of interest. As the

platform passes by the target, the beam direction moves to keep pointing at the target. In this way, $\Delta\theta$ may be made considerably greater than θ_B, and δ_{cr} (spotlight) < δ_{cr} (stripmap). We may wish to make a correction for the variation in received power ($\sim 1/R^4$) as the range to the target varies slightly over the synthetic aperture. This is usually negligible for most spotlight SAR applications but might not be negligible when collection angles are large, such as for the case of foliage-penetration SAR (Section 17.8).

The *synthetic aperture time* t_A required to collect the data for a spotlight SAR image is found as follows:

$$\delta_{cr} \approx \frac{\lambda}{2\Delta\theta} \approx \frac{\lambda R}{2L_{SA}\cos(\theta_{sq})} = \frac{\lambda R}{2Vt_A\cos(\theta_{sq})}$$
$$t_A \approx \frac{\lambda R}{2V\delta_{cr}\cos(\theta_{sq})} \tag{17.6}$$

where V is the platform speed.

Interferometric SAR. Interferometric SAR (InSAR, sometimes also called "IFSAR") refers to the use of two antennas whose signals are combined coherently. InSAR was originally developed by the Jet Propulsion Laboratory (JPL) to detect ocean currents or moving targets.[18,19] The two antennas were displaced horizontally on the platform (along a line parallel to the ground) so that the received echoes from a moving target would be different from the corresponding echoes from fixed targets, and thus movers could be detected and analyzed. Later researchers (e.g., Adams et al.[20]) used two antennas displaced vertically on the platform so that the received echoes from a target above the surface (assumed flat) would be different from the corresponding echoes from a target on the surface, and thus target height could be estimated. Both types of InSAR are discussed in Section 17.8. The former is discussed in "Interferometric SAR for Moving Target Indication (MTI)," and the latter is discussed in "Interferometric SAR for Target Height Measurement."

Inverse SAR. Skolnik[21] presents a discussion of Inverse SAR (ISAR). He states, "In SAR, the target is assumed stationary and the radar is in motion. In ISAR, the target motion provides the changes in relative velocity that cause different doppler shifts to occur across the target" (pp. 375–380). Skolnik also includes a discussion of ISAR images of ships obtained by the U. S. Naval Research Laboratory (NRL) (Musman et al.[22]). An airborne radar obtains a series of images of a ship that is experiencing pitch/roll/yaw in the waves, and the user is able to identify the ship type and characteristics. Musman et al. discuss feature extraction, multiframe processing, and a capability for automatic target recognition (ATR) of ships. ISAR is also widely used for diagnostic measurements on indoor and outdoor radar cross section (RCS) ranges (Knott et al.,[26] p. 516).

Improvements in SAR Resolution. The following example illustrates how SAR crossrange resolution has improved as the preceding types of airborne mapping radar were developed. Let us assume that $\lambda = 0.03$ m (X band), D = aperture diameter = 2 m, $R = 100$ km, $\theta_{sq} = 0$, $V = 180$ m/sec, $L_{DBS} = 10$ m (corresponding to an angular scan rate of 15° sec), and spotlight synthetic aperture length = 5 km ($\Delta\theta \approx 3$ deg).

Then crossrange resolution is approximately as follows for the different modes discussed previously:

SLAR: 1500 m	DBS: 150 m	Unfocused stripmap SAR: 27 m
Stripmap SAR: 1 m	Spotlight SAR: 0.3 m	

17.4 SAR RESOLUTION

In this section, we discuss SAR resolution in more detail. By "resolution," in keeping with common usage, we mean the precision to which we can measure the location of a point target and not necessarily the ability to resolve two point targets. (For a discussion of this issue, see Wolfe and Zissis.[23]) How we do this will be discussed in more detail below.

Since fine range resolution is typically achieved with a single pulse, the corresponding processing is often termed *fast-time processing*. On the other hand, fine crossrange resolution requires multiple pulses, and the corresponding processing is, therefore, often termed *slow-time processing* (Carrara,[3] Richards,[24] and Klemm[25]).

Range Resolution. Strictly speaking, SAR refers to a method for improving crossrange resolution, not range resolution. However, because fine range resolution is so necessary to a successful SAR, and also because of an analogy between range and crossrange resolution, we discuss range resolution briefly here.

Fine range resolution is achieved by transmitting and receiving radar waves characterized by a fairly wide bandwidth B. As an example, consider a carrier frequency $f = 10$ GHz and a 10% bandwidth $B = 1$ GHz. One way to achieve this bandwidth (not the best way, in general, but a way that is easy to describe) is to use a *step-frequency* waveform consisting of a series of "single-frequency" pulses, each with a frequency somewhat greater than that of the previous pulse. (By a "single-frequency" pulse, we mean a pulse consisting of a pure sinusoidal tone multiplied by a rectangular function of duration ("width") τ much longer than the period of the tone. Such a pulse is not truly single-frequency but has a bandwidth of $\approx 1/\tau$. For example, if $\tau =$ 1 microsecond, $B(pulse) = 1$ MHz. This is far less than the bandwidth of the overall step-frequency waveform, equal to the difference between the frequencies of the last and first pulses in the sequence, which is typically hundreds of MHz.)

Step-frequency pulse compression has not proven successful in airborne or spaceborne applications, as compared to the linear-FM (LFM) waveform (see Section 2.6 of Carrara[3]), and is seldom used. Linear-FM (LFM) is a common waveform used in operational high-power radars, because (1) each pulse contains the full bandwidth, thereby enabling the full bandwidth to be transmitted and received much more quickly than with step-frequency—a great advantage when the radar is moving, such as for SAR—and (2) the hardware is relatively inexpensive and mature. LFM has been successfully employed operationally (e.g., in the U. S. Navy AN/APS-116 and AN/APS-137) since the 1970s. The only apparently successful application of step-frequency has been in ground-based instrumentation radars, where it is less expensive to implement and more time is available for data collection (Knott et al.,[26] pp. 35, 540).

Nevertheless, for the moment, we assume the step-frequency waveform because it provides a much simpler example for explaining the principle of range resolution.

Let us assume that the radar transmits a step-frequency waveform consisting of identical groups of pulses, each group consisting of N ($\gg 1$) "single-frequency" pulses of width τ. Within a pulse group, the frequency of a pulse is Δf greater than that of the previous pulse, and the radar transmits (PRF)/N groups per second, where PRF is the pulse repetition frequency. The waveform bandwidth is $(N-1)\,\Delta f \gg 1/\tau$. The phase and amplitude of each pulse echo is digitally recorded by the radar. As shown in Figure 17.1*a*, a discrete Fourier transform (DFT)—typically a fast-Fourier transform (FFT)[27]—is applied to this set of N complex samples of frequency-domain data, resulting in a set of N complex numbers in the time-domain corresponding to echoes (magnitude and phase) that would be returned from a very short pulse of width $\approx 1/B$, sampled at time intervals $\Delta t = 1/B$. This is a simple example of pulse compression. Since an incremental delay Δt corresponds to an incremental downrange distance $\Delta r = c\Delta t/2$, we multiply the DFT output by $c/2$ and obtain the echoes corresponding to a set of downrange distances separated by a pixel width of $c\Delta t/2 = c/2\mathrm{B}$. Thus, the range resolution (strictly speaking, pixel separation) of a step-frequency waveform of bandwidth B is

$$\delta_r = c/2B \tag{17.7}$$

Although it is beyond the scope of this chapter, it can be shown that such a range resolution of $\approx c/2B$ may be obtained using a wide variety of waveform types, as long as the overall transmitted bandwidth is B. For example, Section 7.2 of Sullivan[1] shows that this is true for the LFM waveform.

Crossrange Resolution. Let us now assume that the airborne (or spaceborne) SAR is observing a scene on the Earth's surface consisting of a few point targets and that it transmits and receives N identical pulses each of bandwidth B (presumably, though not necessarily, via the LFM waveform) and determines the downrange position of each

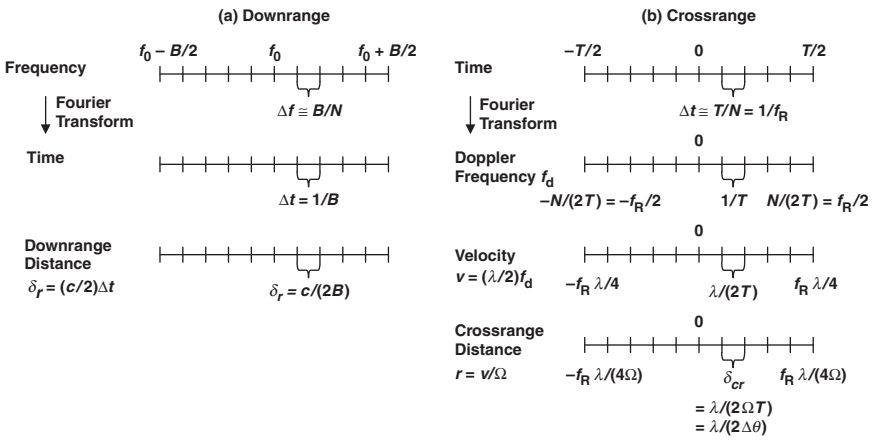


FIGURE 17.1 Range and crossrange resolution: Similar discrete Fourier transform (DFT) processes may be used to obtain range and crossrange resolution.

target with a range resolution of $c/2B$. Let us also assume that the SAR is moving in a straight line at a constant altitude H and constant speed V for a time T along a direction normal to the LOS. The synthetic aperture $L_{SA} = VT$ is assumed small compared with the range R to the center of the target region. As viewed from the target region (also assumed small compared to R), the synthetic aperture subtends the synthetic aperture angle $\approx L_{SA}/R = VT/R$. As the radar moves through its synthetic aperture, it views the scene from slightly different angles. For simplicity, let us assume that during this time the targets do not leave their "range bins" of width $c/2B$. (This assumption will be discussed in Section 17.5, "Range Migration.")

From the viewpoint of the SAR, the scene appears to be rotating with angular velocity $\Omega = V/R$. During the data collection, the total angle through which the scene appears to rotate is $\Delta\theta = \Omega T = VT/R$. A specific point target appears to have a LOS velocity of Ωr relative to the SAR, where r is the crossrange distance of the target from the LOS. These apparent LOS velocities will result in doppler frequencies (in absolute value) of $2v(\text{apparent})/\lambda = 2\Omega r/\lambda$, where λ is the wavelength corresponding to the carrier frequency.

For each range bin, we now have N complex numbers corresponding to different radar echoes in the time domain. As shown in Figure 17.1*b*, these N time-domain echoes may be processed using a DFT to produce a set of N frequency-domain returns. The frequency interval between successive returns is $\Delta f = 1/T$, and the overall frequency interval is $(N-1)/T \approx N/T = \text{PRF} = f_R$. We assume operation at baseband, and thus the frequency in question is the apparent doppler frequency of the targets. We convert this to crossrange by multiplying by $\lambda/2\Omega = \lambda R/2V$:

$$\delta_{cr} \approx \frac{\lambda}{2\Omega T} = \frac{\lambda}{2\Delta\theta} \approx \frac{\lambda R}{2L_{SA}} = \frac{\lambda R}{2VT} \tag{17.8}$$

Note that our assumption that $R \gg VT$ permits the use of small-angle approximations for small $\Delta\theta$. When $\Delta\theta$ is not small, Eq. 17.8 must be appropriately modified.

Summary of SAR Resolution. We have, therefore, now derived the two basic formulas for SAR resolution:

$$\begin{aligned} \delta_r &= c/2B \qquad \text{Range Resolution} \\ \delta_{cr} &= \frac{\lambda}{2\Delta\theta} \approx \frac{\lambda R}{2L_{SA}} \quad \text{Crossrange Resolution} \end{aligned} \tag{17.9}$$

The processing described produces a two-dimensional array of complex numbers, i.e., each consists of a magnitude and a phase. This ordered array of complex numbers as a function of downrange and crossrange produces a complex radar *image,* i.e., each pixel consists of a magnitude and a phase. Typically the magnitude squared (representing pixel energy) is displayed.

As shown in Chapter 6 of Sullivan,[1] a point target in the scene transforms to a two-dimensional *point-spread function* (PSF) in the radar image, so-called because a point target is displayed in the image as somewhat "spread out." This PSF is characterized by a *mainlobe* and *sidelobes* in both range and crossrange. Usually *weighting* (also called *tapering* or *apodization*) is applied in the processing, resulting in considerably lower sidelobes, but at the expense of a somewhat broader mainlobe, a price that the user is typically willing to pay. Many types of weighting exist. If no weighting is applied, the PSF is a $(sinc)^2$ function of the form $(\sin(x)/x)^2$. In that case, the formulas above for SAR resolution

represent the distance from the mainlobe peak to the first null. Note that this definition of SAR resolution is different from the more common definition of the half-power width of the mainlobe. With no weighting, the latter is (0.886) times the values given above; thus the two definitions are not very different. We prefer the former definition because, for no weighting, it results in a simpler formula without the introduction of the factor of (0.886). (More details concerning weighting are given in Section 17.6.)

There are at least two mathematically equivalent ways of considering SAR. As we have developed the concept so far, the crossrange resolution may be considered to result from the doppler shifts resulting from the different apparent line-of-sight velocities of different parts of the scene. However, the crossrange resolution may also be considered to result from the large synthetic aperture, much as fine crossrange resolution would also result from a large real aperture. Per Eq. 17.8, the crossrange resolution (peak-to-first null) of a SAR is finer by a factor of two than the resolution for a RAR of equal aperture. An intuitive explanation for this interesting result is that for RAR, the echo received at a particular aperture location results from energy transmitted from *all locations* in the aperture, whereas for SAR, the echo received at a particular aperture location results from energy transmitted from *that (known) location* in the aperture; i.e., more information is received (Carrara et al.,[3] p. 36). Stimson[5] provides a detailed explanation for this result (pp. 416–417).

Figure 17.2 presents a comparison of RAR and SAR.

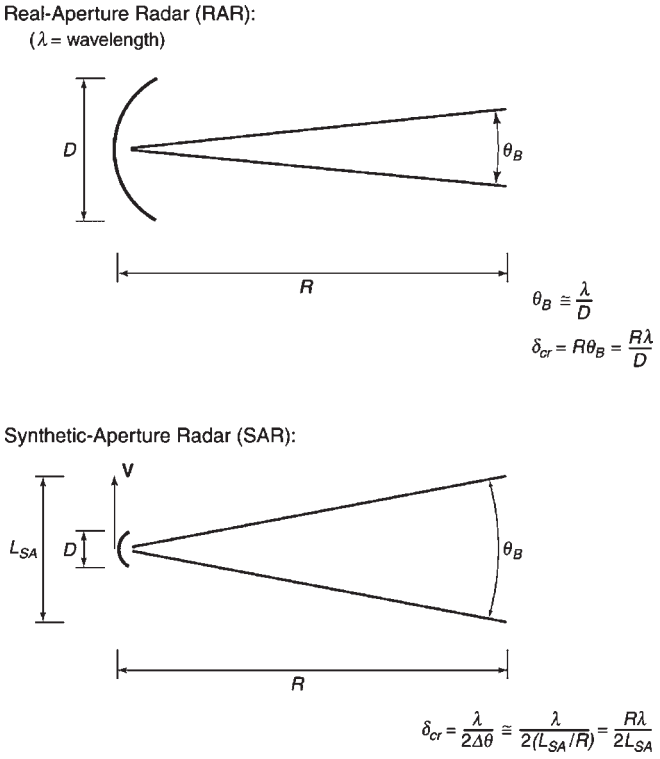


FIGURE 17.2 Comparison of RAR and SAR: The crossrange resolution (peak-to-first-null of antenna pattern) of a SAR is one-half that of a real aperture radar (RAR) of the same aperture diameter. (*Courtesy of SciTech Publishing, Inc.*)

The sidelobes are also different for RAR and SAR. For a RAR with an unweighted aperture function, the transmitted intensity at the peak of the first angular sidelobe is reduced by –13 dB, and the received intensity from a target in that direction is thus reduced by –26 dB. For a SAR, typically the entire region of the image is in the main beam during the data collection, and there are no effects from the sidelobes of the physical antenna. The sidelobes result solely from the processing, and (with no weighting) the first sidelobe is reduced by –13 dB relative to the main beam.

17.5 KEY ASPECTS OF SAR

Range and Velocity Contours. Using fine range resolution, a radar can distinguish between targets at different ranges. A particular target may be determined to be located on a constant-range-contour. In 3D space, these contours are concentric spheres with the radar at the center (Figure 17.3*a*).

Similarly, using doppler processing, the radar can distinguish between targets of different apparent velocities. If **V** is the platform velocity and θ is the angle between **V** and the LOS to a stationary target, then the apparent LOS speed of the target is $V_{\text{LOS}} = -V\cos\theta$ (Figure 17.4*b*). In 3D space, surfaces of constant V_{LOS} are circular cones with axis **V** and generating angle θ, with the radar at the vertex. The negative sign occurs

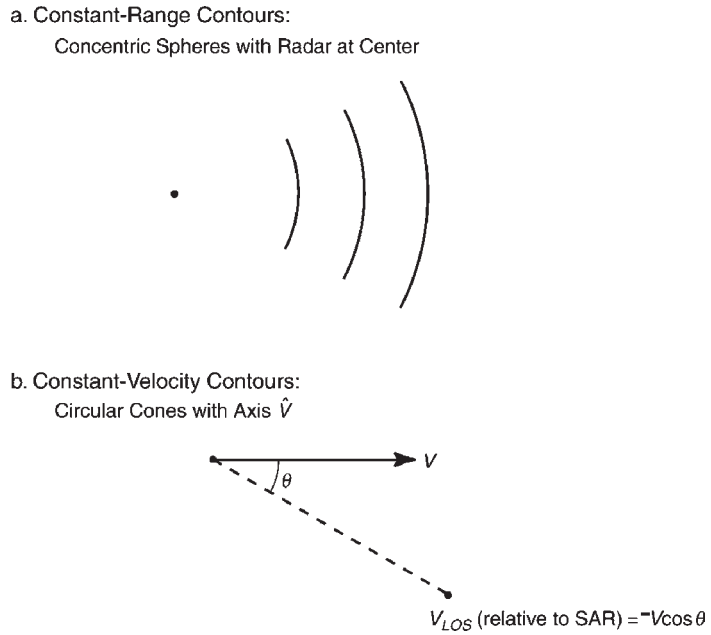


FIGURE 17.3 Range and velocity contours in 3D space: (*a*) Constant-range contours are concentric spheres with the radar at the center. (*b*) Constant-apparent-velocity contours are circular cones with the radar at the vertex and the axis along the platform velocity vector. (*Courtesy of SciTech Publishing, Inc.*)

because we define LOS velocity as dR/dt, where R is the range to the target. Thus, a positive dR/dt corresponds to a receding target and results in a negative doppler frequency shift, and a negative dR/dt corresponds to an incoming target and results in a positive doppler frequency shift.

Consider an airborne radar with an isotropic antenna pattern, with constant velocity along a straight line parallel to a flat ground[3] (Figure 17.4*a*). On the ground, the constant-range contours are the intersections of concentric spheres with the ground—a set of concentric circles with the subradar point at the center (Figure 17.4*b*). The contours of constant LOS velocity (called *isodops*) correspond to the intersections between the set of cones and the flat ground—a set of nested hyperbolas (Figure 17.4*c*). Figure 17.4*d* shows the combination of constant-range circles and the *isodops*. The "nadir line" shown in Figure 17.4 is the locus of subradar points as the platform flies along its path.

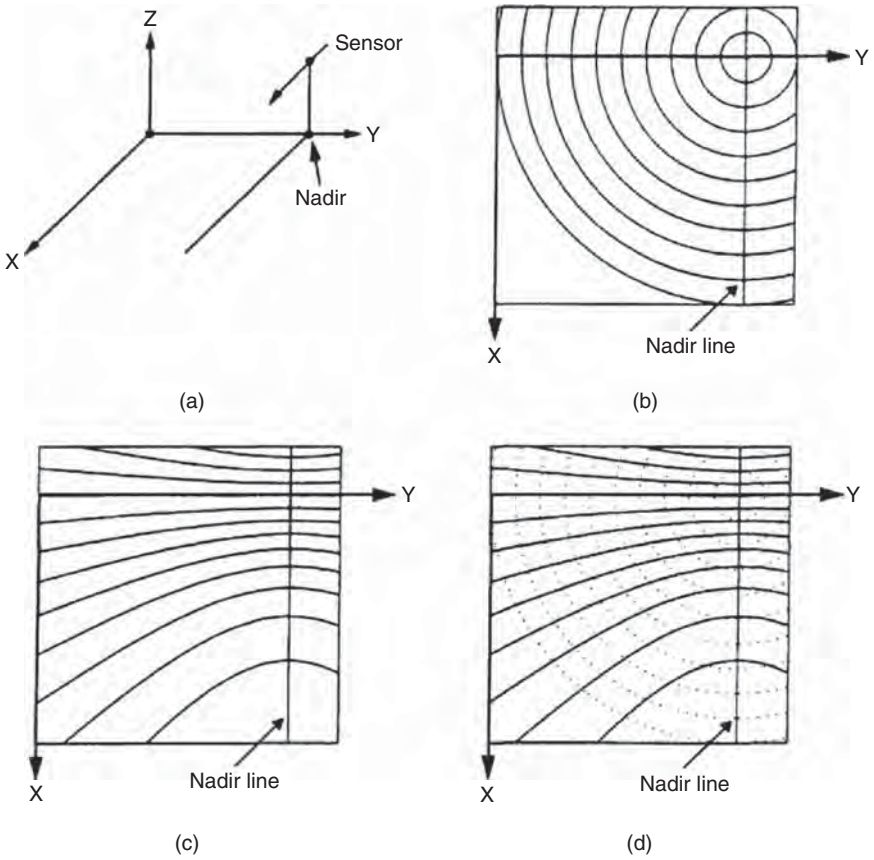


FIGURE 17.4 Range and velocity contours on the Earth's surface: (*a*) Collection geometry. (*b*) Constant-range contours are concentric circles with the subradar point at the center; "nadir line" refers to the locus of subradar points. (*c*) Constant-apparent-velocity contours ("isodops") are confocal hyperbolas with the axis parallel to the platform velocity vector. (*d*) Set of intersecting concentric circles and confocal hyperbolas. (*Courtesy of SciTech Publishing, Inc.*)

Through appropriate range-doppler processing, returns from each intersection cell may be distinguished. Over a small angle about the broadside direction, the range contours and the isodops are essentially orthogonal to each other. The resulting radar returns may be displayed to yield an image of the ground. At a nonzero squint angle, the isodops are not orthogonal to the range contours; however, additional processing corrections can still usually result in an essentially undistorted ground image.

Motion Compensation. The basic theory of SAR relies on the assumption that the platform, and therefore the SAR antenna, is traveling along a straight-line flight path at constant velocity parallel to the ground at constant altitude. This is not exactly true, and for successful SAR imaging, it is necessary that the deviations of the antenna from this nominal flight path be measured, recorded, and compensated for in the processing. This procedure is known as *motion compensation* (sometimes abbreviated *mocomp*). For example, at a particular moment, as a particular frequency is transmitted, if the antenna is estimated to have deviated a distance d away from the nominal flight path along the LOS, the phase correction

$$\Delta\phi = \frac{4\pi d}{\lambda} = \frac{4\pi df}{c} \tag{17.10}$$

with appropriate sign, is added to the measured phase at the frequency f to produce the best estimate of what the recorded phase would have been if the platform had not deviated from the nominal flight path. Similarly, if the platform speed is not constant, the received data are interpolated to produce the best estimate of what they would have been if the speed had been constant.

When the platform is an aircraft, an on-board *Inertial Navigation System (INS)* uses accelerometers and gyroscopes to measure the deviations. Sometimes a smaller *Inertial Measurement Unit* (IMU) relying on the same general principles is "strapped down" very near the antenna. Without an absolute reference frame, the outputs of any INS or IMU will drift with time as errors accumulate. An absolute frame for position and velocity may be obtained from the *global positioning system (GPS)*, a constellation of at least 24 satellites in polar Earth orbit providing continuous reference signals for determination of precise position and velocity.[28]

Slant and Ground Planes. When a SAR image is initially produced, the range pixel size δ_r is usually a constant. (It is generally chosen to be somewhat less than $c/2B$, e.g., 0.75 ($c/2B$), to ensure adequate sampling). As illustrated in Figure 17.5, the actual ground locations that correspond to these range samples are not spaced at constant intervals in ground range. Near the scene center, they are spaced at

$$\delta_g \cong \delta_r / \cos\ \psi \tag{17.11}$$

where ψ is the grazing angle. At ground ranges closer to the radar, they are spaced still farther apart because of the spherical range contours. Near the scene center, the image corresponds to the projection of the ground onto a *slant plane*; this plane is determined by the LOS and its perpendicular in the ground plane. We often refer to this type of image as a *slant-plane image*. By appropriate interpolation and resampling, a *ground-plane image* with δ_g = constant = δ_{cr} may be produced. Ground-plane imagery with minimal distortion is necessary if comparison is to be made with maps or with imaging taken from other sensors, such as optical sensors or other SARs.

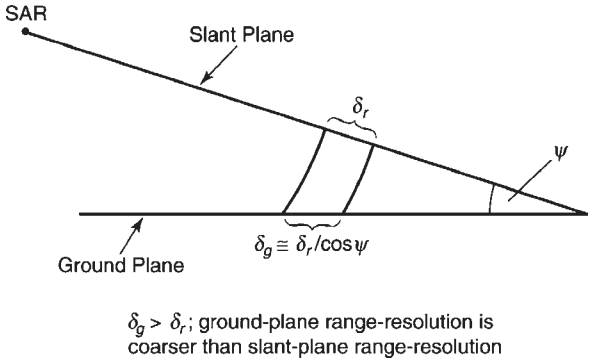


FIGURE 17.5 Slant and ground planes: The slant plane is determined by the radar line-of-sight and its perpendicular in the ground plane. Ground-plane range resolution is coarser than slant-plane range resolution. (*Courtesy of SciTech Publishing, Inc.*)

Pulse Repetition Frequency (PRF) Requirements for SAR. For broadside operation, the apparent angular velocity of the scene rotation is

$$\Omega = \frac{V}{R} \tag{17.12}$$

With respect to the radar, the relative velocity of point *A* (see Figure 17.6) at the first null of the main beam is

$$v_A = -\Omega r = -\left(\frac{V}{R}\right)\left(\frac{\lambda R}{2D}\right) = -\frac{\lambda V}{2D} \tag{17.13}$$

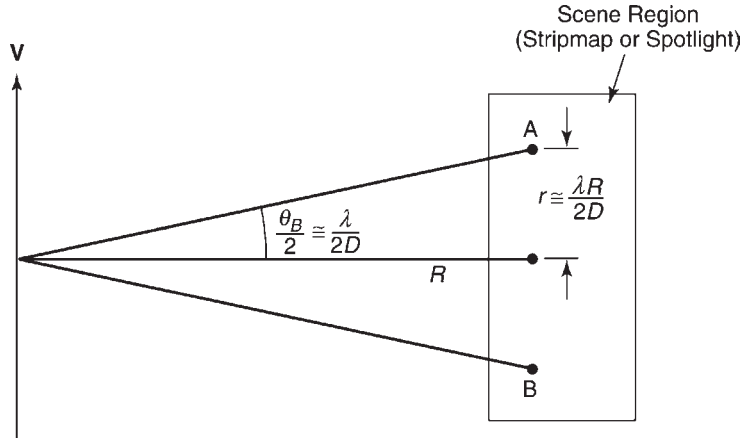


FIGURE 17.6 Minimum PRF for SAR: The apparent line-of-sight velocity of point A is toward the radar, whereas the apparent line-of-sight velocity of point B is away from the radar. These determine the minimum PRF of $2V/D$, where V = platform speed and D = antenna physical diameter. (*Courtesy of SciTech Publishing, Inc.*)

Similarly, the relative velocity of point B on the opposite side of the beam is

$$v_B = \Omega r = \left(\frac{V}{R}\right)\left(\frac{\lambda R}{2D}\right) = \frac{\lambda V}{2D} \tag{17.14}$$

Thus, the range of relative velocities in the scene is

$$\Delta v = \Omega r - (-\Omega r) = 2\Omega r = \frac{\lambda V}{D} \tag{17.15}$$

The range of doppler frequencies received from the scene is

$$\Delta f_d = \frac{2}{\lambda}\frac{\lambda V}{D} = \frac{2V}{D} \tag{17.16}$$

Thus, to avoid velocity ambiguity, the PRF must be at least $2V/D$. Writing

$$f_R(\min) = \frac{2V}{D} = \frac{1}{t_R(\max)} \tag{17.17}$$

we have

$$Vt_R(\max) = \frac{D}{2} \tag{17.18}$$

Thus, the distance traveled by the platform during the time between pulses (t_R) must be no more than $D/2$, and the SAR must transmit at least two pulses as its physical antenna passes a stationary point in space.

We also frequently want range-unambiguous operation, which implies

$$\frac{2V}{D} \le f_R < \frac{c}{2R} \tag{17.19}$$

For example, if $V = 180$ m/s (≈ 350 kts), $D = 2$ m, and $R = 150$ km, then 180 Hz < f_R < 1000 Hz.

Eq. 17.19 is really two equations. We use < for the second equation instead of ≤ because when the equality is used, a pulse is being transmitted just as the echo from the previous pulse is being received, resulting in eclipsing and consequent loss of the received information.

Skolnik[21](pp. 520–521) points out that, since for stripmap SAR the crossrange resolution $\delta_{\text{cr}} \cong D/2$ (Eq. 17.4), Eq. 17.19 becomes

$$\frac{V}{\delta_{\text{cr}}} \le f_R < \frac{c}{2R} \tag{17.20}$$

This leads to the condition

$$\frac{R}{\delta_{\text{cr}}} \le \frac{c}{2V} \tag{17.21}$$

Thus, the unambiguous range R_u and the resolution of a stripmap SAR cannot be selected independently of one another. Skolnik further points out, quoting Bayma and McInnes,[29] that more sophisticated reasoning leads to the condition

$$\frac{R_u}{\delta_{\text{cr}}} \le \frac{c}{4.7V} \tag{17.22}$$

Skolnik[21] continues, "When a SAR images the ground from an elevated platform, the unambiguous range can correspond to the distance between the forward edge and the far edge of the region to be mapped. This requires that the elevation beamwidth be tailored to illuminate only the swath S that is to be imaged by the radar. The swath S is often much smaller than the maximum range so that the PRF can be increased to allow the unambiguous range R_u to encompass the distance $S \cos \psi$, where ψ is the grazing angle" (p. 521). For a stripmap SAR, Eq. 17.22 then becomes

$$\frac{S}{\delta_{\text{cr}}} \leq \frac{c}{(4.7)V \cos\psi} \tag{17.23}$$

Range Migration. As we have seen, a SAR may obtain range resolution of $\delta_r = c/2B$, where B = signal bandwidth, and obtain cross-range resolution through doppler processing, with $\delta_{\text{cr}} = \lambda/2\Delta\theta$. If we wish to prevent *range migration* (movement of a point target from one range bin to the next during the time required to collect the data for image formation), we would require that ΔR, the variation of range during the data collection (over the synthetic aperture), be less than δ_r.

We consider the formation of a SAR image after collecting data over an aperture time t_A. The SAR flight path is a straight line at constant speed and altitude over a flat ground. From Levanon[17] we have

$$R_{\max} \cong R_0 + \frac{V^2(t_A/2)^2}{2R_0}, \qquad R_{\min} = R_0 \tag{17.24}$$

where R_0 = distance from the radar to the scene center in the middle of the data collection interval, and $R_{\max}$ = distance from the radar to the scene center at the beginning and end of the data collection interval. Then

$$\Delta R = R_{\max} - R_{\min} = \frac{(Vt_A)^2}{8R_0} = \frac{L_{\text{SA}}^2}{8R_0} = \frac{R_0(\Delta\theta)^2}{8} = \frac{R_0\lambda^2}{32\delta_{\text{cr}}^2} < \delta_r \tag{17.25}$$

The last inequality is necessary for the condition of no range migration. For example, parameters for a SAR might be $R_0 = 200$ km, $\lambda = 0.03$ m, and $\delta_r = \delta_{\text{cr}} = 1$ m; then, $\Delta R = 5.6$ m $> \delta_r$, and the condition is not satisfied. Thus, *the processor must usually correct for range migration*. However, this is typically accomplished by modern SAR processing methods.[3] For spotlight SAR, the *polar format algorithm*[3,11] is often used to accomplish this correction.

Other Processing Functions. Curlander et al.[4] provide a detailed discussion of several key options for SAR processing beyond simple image formation.

- *Clutterlock* (Curlander et al.,[4] Chapter 5) refers to the use of information in the received signals to ascertain the center frequency of the echoes from the ground (clutter) and compensate for sideways drift of the platform.
- *Autofocus* (Curlander et al.,[4] Chapter 5) describes the use of information in the (complex) image itself to estimate and correct phase errors, and then reprocess and sharpen the image (see also Carrara et al.[3]).
- *Calibration* (Curlander et al.,[4] Chapter 7) refers to the use of targets of known radar cross section (RCS) in the scene to obtain the absolute level of RCS per pixel and thus σ^0, the ground RCS per unit area.
- *Geolocation* (Curlander et al.,[4] Chapter 8) is the process of determining the absolute latitude and longitude of pixels in the SAR image, typically using information from the GPS.

17.6 SAR IMAGE QUALITY

It is clearly important for a SAR to produce high-quality imagery. Image quality is typically measured using several *image-quality metrics* (IQMs) described in the following sections. More detailed discussion of SAR imagery is given by Henderson and Lewis[30] and Oliver and Quegan.[31]

Point-Spread Function (PSF). A point target may be considered an impulse input to a SAR processor, and the PSF in the image may be regarded as an impulse response (IPR). The primary IQM for most SARs is the width (meters) of the PSF mainlobe at its "half-power points," or points where the intensity (power, proportional to voltage squared) is one half, or −3 dB, relative to the mainlobe peak. This IQM is typically referred to as the "3-dB width."

To obtain fine resolution in either range or crossrange, a Fourier transform (FT) is performed on a set of collected data. Since all data samples have essentially the same magnitude, for either range or crossrange we essentially perform a FT on a rectangular function, which produces a $(sinc)^2$ function $(sin(x)/x)^2$, with a 3-dB width of $(0.886)\delta_{pn}$ (where δ_{pn} = peak-to-first-null interval) and a first sidelobe –13.3 dB below the peak.

As mentioned in Section 17.4, when a tapering, or weighting, function multiplies the rectangular input, the result is typically a function with broader mainlobe and lower sidelobes than the $(sinc)^2$ (see Section 2.2.2 of Sullivan[1]). A typical weighting function used in SAR processing is Taylor weighting, with the first sidelobe constrained to be –35 dB below the peak and "nbar = 5" (see Section D.2 of Carrara et al[3]), which produces a widened main beam of 3-dB IPR value $(1.19)\delta_{pn}$. Another choice is Hann[17] (or "Hanning"[3]) weighting, which results in an even wider main beam of $(1.43)\delta_{pn}$; the first sidelobe is –31.7 dB below the peak, and the far sidelobes are very low compared with uniform or Taylor weighting. An excellent discussion of over 20 weighting functions (not including Taylor) is given by Harris.[32]

Signal-To-Noise Ratio (SNR). For a real radar, for which the exact phase of the target echo can never be known in advance, the greatest achievable signal-to-noise ratio (SNR) is (Sullivan,[1] Section 4.2.2)

$$\text{SNR} = \frac{E}{kT_0F} \tag{17.26}$$

$$E = P_{\text{Rx–avg}}t_A \tag{17.27}$$

where E = collected energy; k = Boltzmann's constant = 1.38×10^{-23} joule/Kelvin; T_0 = standard temperature (290 K); F = "noise figure," which is typically about 2; $P_{\text{Rx–avg}}$ = average received power; and t_A = time to form the synthetic aperture. (The denominator of Eq. 17.26 is correct only if the temperature of the radar is the same as that of the scene, which we assume; see Chapter 2 and Section 1.11 of Sullivan[1]):

$$E = \frac{P_{\text{Tx-avg}}G^2\lambda^2\sigma}{(4\pi)^3R^4(\text{Loss})} \cdot t_A = \frac{P_{\text{Tx-avg}}A^2\eta^2\sigma}{4\pi R^4\lambda^2(\text{Loss})} \cdot t_A \tag{17.28}$$

where $P_{\text{Tx-avg}}$ = average transmitted power, G = antenna gain, σ = target radar cross section (RCS), A = antenna aperture area, η = antenna efficiency, and the radar losses are represented by (*Loss*).

For SAR, at $\theta_{sq} = 0$, from Eq. 17.6,

$$t_A = \frac{\lambda R}{2V\delta_{\text{cr}}} \tag{17.29}$$

Thus

$$\text{SNR} = \frac{P_{\text{Tx-avg}} G^2 \lambda^3 \sigma}{2(4\pi)^3 R^3 k T_0 F(\text{Loss}) V \delta_{\text{cr}}} = \frac{P_{\text{Tx-avg}} A^2 \eta^2 \sigma}{8\pi R^3 \lambda k T_0 F(\text{Loss}) V \delta_{\text{cr}}} \tag{17.30}$$

If a flat ground is being observed, then

$$\sigma = \sigma^0 \delta_{\text{cr}} \delta_r / \cos\psi \tag{17.31}$$

where σ^0 characterizes the ground RCS per unit area and δ_r = pixel width in slant range. Then the signal-to-noise ratio (SNR) is

$$\text{SNR} = \frac{P_{\text{Tx-avg}} G^2 \lambda^3 \sigma^0 \delta_r}{2(4\pi)^3 R^3 k T_0 F(\text{Loss}) V \cos\psi} = \frac{P_{\text{Tx-avg}} A^2 \eta^2 \sigma^0 \delta_r}{8\pi R^3 \lambda k T_0 F(\text{Loss}) V \cos\psi} \tag{17.32}$$

This agrees with Skolnik,[21] Eq. 14.15, and Curlander et al.,[4] Eq. 2.88.

It is very useful to consider the *noise-equivalent sigma-zero* ($NE\sigma^0$), defined as the level of σ^0 that produces a received power equal to the thermal noise power, i.e., an SNR of unity. We set SNR = 1 and have

$$NE\sigma^0 = \frac{2(4\pi)^3 R^3 k T_0 F(\text{Loss}) V \cos\psi}{P_{\text{Tx-avg}} G^2 \lambda^3 \delta_r} = \frac{8\pi R^3 \lambda k T_0 F(\text{Loss}) V \cos\psi}{P_{\text{Tx-avg}} A^2 \eta^2 \delta_r} \tag{17.33}$$

For example, if $R = 200$ km, $T_0 = 290$ K, $F = 2$, Loss = 5, $V = 180$ m/s, $\psi = 10^\circ$, $P_{\text{avg}} = 700$ W, $G = 34$ dB, $\lambda = 0.03$ m (X band), and $\delta_r = 0.3$ m, then $NE\sigma^0 = -22$ dB.

A clear SAR image must have an SNR greater than about 5 dB. From Barton[33] and summarized in Section 3.2 of Sullivan,[1] we see that this example SAR could image "wooded hills," i.e., $\sigma^0 \approx -17$ dB, with SNR ≈ 5 dB, but could not image "flatland" (perhaps desert) at $\sigma^0 \approx -27$ dB, since SNR ≈ −5 dB.

Integrated Sidelobe Ratio (ISLR). An actual PSF typically resembles the theoretical PSF but is somewhat different from it, especially in the sidelobes, due to phase noise, motion compensation imperfections, and other "real-world" effects. A useful figure of merit is the integrated sidelobe ratio (ISLR), defined as[3]

$$ISLR = \frac{\text{Integral over PSF Sidelobes}}{\text{Integral over PSF Mainlobe}} \tag{17.34}$$

ISLR is usually measured in dB; a typical value might be −20 dB. Low ISLR is clearly desired.

Multiplicative Noise Ratio (MNR). Thermal noise (generally internal when the radar is a SAR) is often referred to as *additive noise*, since it adds to the scene independent of the scene content. Another type of unwanted background in a SAR image

is often called *multiplicative noise* (not really noise in the strict sense), which is proportional to the average scene intensity.

Carrara et al.[3] define multiplicative noise as follows: "The principal contributors to multiplicative noise are the integrated sidelobes of the system impulse response, the energy present in the scene as a result of range and azimuth ambiguities, and digital [i.e., from quantization in the analog-to-digital converter] noise" (p. 332).

The *multiplicative noise ratio (MNR)*[3] of a SAR image is defined as the ratio of the image intensity in *no-return area (NRA)* (not including thermal noise) divided by the average image intensity in a relatively bright surrounding area (in principle not including thermal noise). An NRA is an area with essentially zero return—for example, a shadow area, a very smooth area such as a calm lake, or a specially constructed large sheet of aluminum.

Another, similar SAR image quality metric (IQM) is the *contrast ratio (CR)*, defined as the ratio of the average intensity of a typical bright region in a SAR image to the intensity of an NRA. If thermal noise is negligible, then CR = 1/MNR.

Comparison of SAR Imagery and Optical Imagery. The human eye is, of course, a system for producing images using visible light. The light hits the lens and is focused upon the retina, and the resulting image is transmitted to the brain. Over many millennia, humans have become fully accustomed to seeing and processing this visible imagery. Therefore, upon seeing a SAR image, we may instinctively assume that it has certain properties of a visible image, which, in fact, it does not possess. Optical imagery is based on an "angle-angle" principle, whereas SAR imagery is based on a very different "range-crossrange" principle.

The top illustration in Figure 17.7 illustrates the appearance of a flat landscape to the human eye (or a camera). The terrain is illuminated by sunlight, at least partially diffused through the atmosphere. At the eye, each pixel subtends the same azimuth and elevation angles. Thus, pixels farther from the eye are larger (coarser resolution), in both downrange and crossrange, than pixels closer to the eye.

The bottom illustration in Figure 17.7 shows that, for a SAR image, the situation is quite different (assuming adequate SNR). The range pixel size δ_r is

$$\delta_r \cong \frac{c}{2B} \cdot \frac{1}{\cos \psi} \tag{17.35}$$

Pixels farther from the SAR are smaller in range (smaller grazing angle and finer downrange resolution) than pixels closer to the SAR; and crossrange resolution is independent of range.

When we display a SAR image, especially of a large landscape, it is usually most satisfying to display it with the SAR direction at the top. The finer resolution pixels are at the bottom, just as they are with a naturally oriented optical image. This orientation tends to look most natural to a human observer.

Because SAR imagery and optical imagery are collected using entirely different physical principles, we should not be surprised if they look different. A good example of this is provided in a SAR image of the Washington Monument, courtesy of the Environmental Research Institute of Michigan, now General Dynamics, Ypsilanti, MI. The top illustration in Figure 17.8 illustrates the collection geometry (and the schematic result) of optical imagery of the Washington Monument with the Monument's shadow pointed toward the observer. We assume that the sun is to the south of the Monument and the observer is to the north. The image shows a shadow on the north

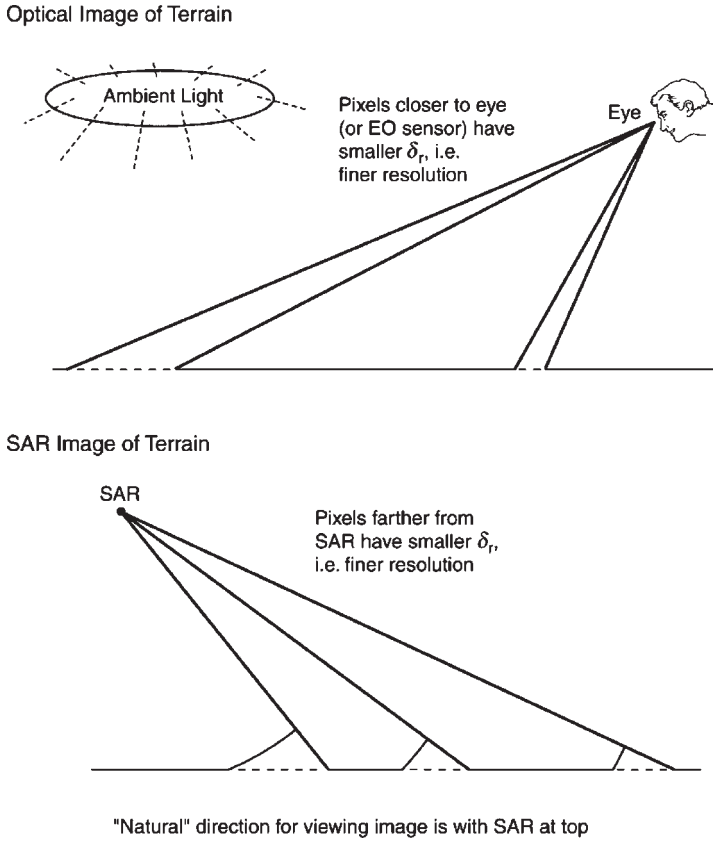


FIGURE 17.7 Comparison of SAR and optical imagery: Optical imagery is based on an "angle-angle" principle. SAR imagery is based on a "range-crossrange" principle. Often optical and SAR images of the same target region do not look the same to a human observer. (*Courtesy of SciTech Publishing, Inc*.)

side cast by the sun. The portion of the Monument visible in the image is the north side, illuminated by diffusely scattered sunlight. In comparison, the bottom illustration in Figure 17.8 shows the geometry and result of SAR imagery, again with the shadow on the north side. This time the shadow is cast by the SAR itself. The portion of the monument visible in the image is the *south* side. Figure 17.9 shows the SAR image. It does not look entirely like an optical image, nor should it.

Another difference between SAR and optical images is the presence of *speckle* (see Section 2-5.1 in Henderson and Lewis[30]) in the former. Let us consider a particular pixel of a complex image of complicated terrain, such as vegetation. (By "pixel" we mean the complex number—magnitude and phase—that, after SAR processing, corresponds to a particular location on the ground.) If only one scatterer were in the region of ground represented by the pixel, then the pixel magnitude and phase would be a function of the scatterer's exact position. Since the region represented by the pixel typically contains many scatterers, the complex pixel value is the sum of the

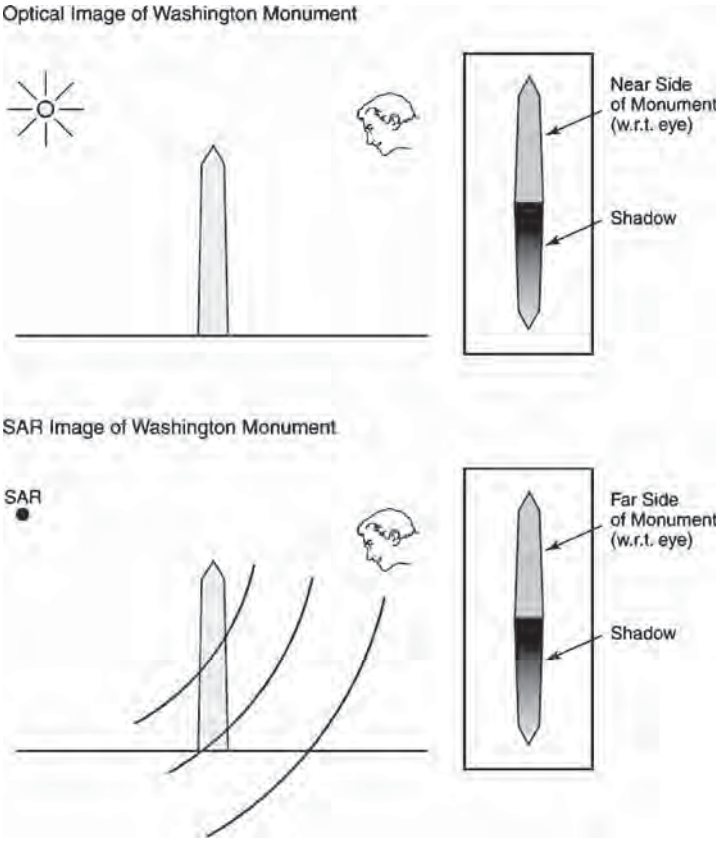


FIGURE 17.8 Principles of imaging the Washington Monument: For the geometry shown, the optical image shows the side of the Monument on the same side as the shadow, whereas the SAR image shows the side of the Monument on the opposite side of the shadow. (*Courtesy of SciTech Publishing, Inc.*)

complex numbers that each results from one of the scatterers. Thus, when terrain, especially vegetation, is imaged, the amplitude (voltage) of a particular pixel is the magnitude of the complex sum of the coherent returns from many scatterers within the pixel. In another nearby pixel, even if the terrain is nominally the same as in the first pixel, the coherent returns will add differently and the pixel magnitude will be somewhat different. This phenomenon, characteristic of coherent imagery, causes SAR imagery of terrain to exhibit more pixel-to-pixel fluctuation (speckle) than corresponding optical imagery.

Stimson[5] points out, "Sometimes the beam of the real antenna may be wide enough to enable the same area to be mapped several times without changing the antenna's look angle. This is called *multilook mapping*. When the maps are superimposed (i.e., when [the magnitudes of] successive returns from each resolution cell are averaged), the effects of scintillation [i.e., speckle] are reduced" (p. 432).

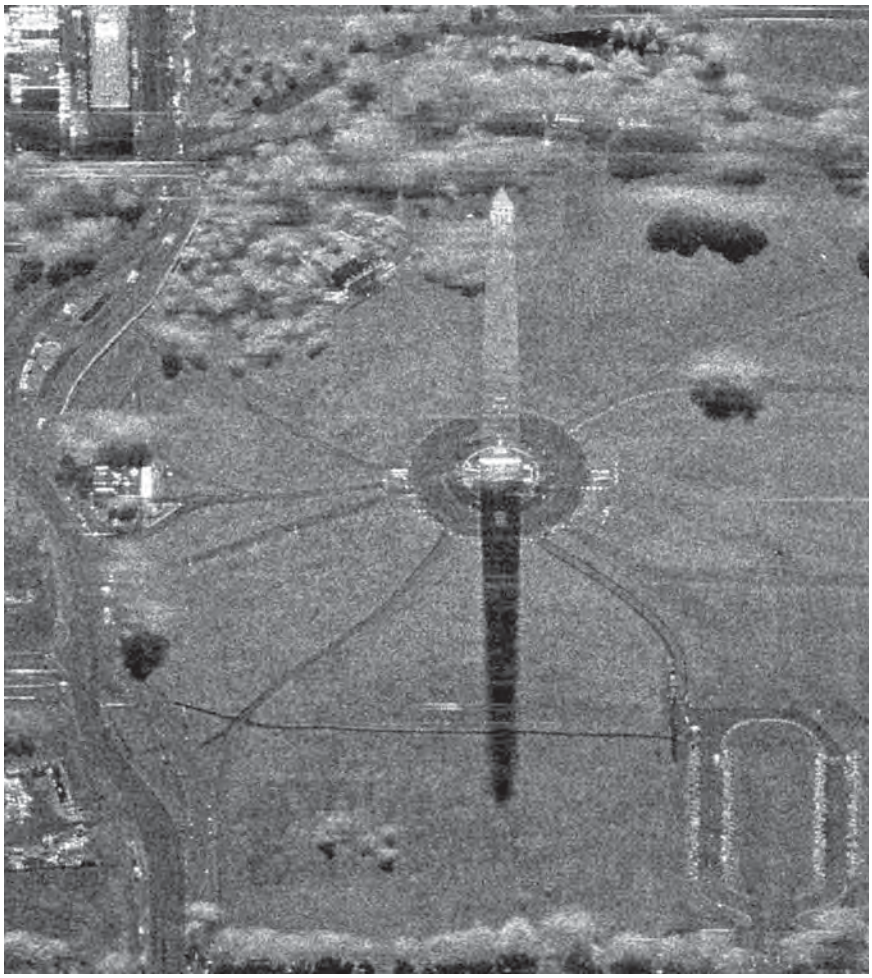


FIGURE 17.9 SAR image of the Washington Monument: The SAR image shows the side of the Monument that is on the opposite side of the shadow, which may appear counterintuitive to a human observer. (*Courtesy of General Dynamics, Ypsilanti, Michigan*)

17.7 SUMMARY OF KEY SAR EQUATIONS

A review of the basic equations of SAR follows:

- Range resolution = $\delta_r \cong c/2B$ (c = speed of light, B = pulse bandwidth)
- Crossrange resolution = $\delta_{cr} \cong \lambda/2\Delta\theta$
  (λ = wavelength, $\Delta\theta$ = angle subtended by synthetic aperture)
- Physical beamwidth $\cong \lambda/D$ (D = antenna diameter)

- Crossrange resolution for stripmap SAR:

$$\delta_{\text{cr}} \approx \frac{\lambda}{2\Delta\theta} \approx \frac{\lambda}{2(\lambda/D)} = \frac{D}{2} \tag{17.36}$$

- Image Collection Time: $t_A = \lambda R/(2V\delta_{\text{cr}}\cos\theta_{\text{sq}})$
- PRF: $f_R = PRF \geq 2V/D$ (V = platform velocity)
- PRF limits for unambiguous range, $f_R < c/2R$:

$$\frac{2V}{D} \leq f_R < \frac{c}{2R} \tag{17.37}$$

$$\text{For stripmap SAR,} \qquad \frac{S}{\delta_{\text{cr}}} \leq \frac{c}{(4.7)V\cos\psi} \tag{17.38}$$

- Signal-to-noise ratio:

$$SNR = \frac{P_{\text{Tx-avg}}G^2\lambda^3\sigma^0\delta_r}{2(4\pi)^3R^3kT_0F(\text{Loss})V\cos\psi} = \frac{P_{\text{Tx-avg}}A^2\eta^2\sigma^0\delta_r}{8\pi R^3\lambda kT_0F(\text{Loss})V\cos\psi} \tag{17.39}$$

17.8 SPECIAL SAR APPLICATIONS

In this section, we briefly discuss several specific aspects of SAR, specifically polarimetric SAR, moving targets, vibrating targets, measurement of object height, and foliage-penetration SAR.

Polarimetric SAR. Usually when a radar transmits a pulse at a particular polarization (e.g., horizontal—H), it receives the echoes at the same polarization. Some radars are capable of transmitting at one polarization and receiving at two orthogonal polarizations (e.g., horizontal [H] and vertical [V] or right-circular [R] and left-circular [L]). Furthermore, some radars can transmit at either of two orthogonal polarizations and receive at either of the transmitted polarizations; and the choice of transmitted and received polarizations can be varied from pulse to pulse. If the phases as well as the magnitudes of the echoes are obtained, then such a radar is fully *polarimetric*. We may designate the choice of polarizations as follows: HV is "transmit H, receive V," and so forth.

Fully polarimetric SARs have been demonstrated (Sullivan et al.[34] and Held et al.[35]). For example, Sullivan et al.[34] includes HH and HV X-band SAR images taken of the same scene at the same time using HH and HV modes interleaved on a pulse-to-pulse basis. The SAR described could, using pulse-to-pulse interleaving, transmit/receive first HH, then HV, then VH, and then VV data, and produce four corresponding simultaneously collected complex images, such that the phases of the corresponding pixels in the four images bear a specific relationship to each other, depending on the target type.

Novak et al.[36] developed an optimal *polarimetric whitening filter* for enhanced target detection in such sets of fully-polarimetric SAR images. Using fully polarimetric data from a 33-GHz SAR, they showed that dihedral reflectors look quite different from trihedral reflectors in fully-polarimetric SAR imagery. Few dihedrals exist in K_a-band natural clutter; thus, if a portion of a SAR image corresponds closely to a dihedral, then the region is likely to contain cultural (human-made) object(s).[37]

Moving Targets in a SAR Image.

Displacement of a Moving Target. The basic theory of SAR assumes that the ground (scene) is stationary. A moving target in the scene will have a "wrong" relationship between its location and its line-of-sight velocity. If the target motion is in a straight line at constant speed, the target image will be displaced in crossrange by

$$r_{\text{displ}} = \frac{V_{\text{LOS}}}{\Omega} = \frac{V_{\text{LOS}} R}{V} \tag{17.40}$$

where Ω is the apparent rotation rate of the scene relative to the radar and V_{LOS} is the target velocity component along the radar line-of-sight. In general, complicated motion of a target during a SAR data collection prevents formation of a clear SAR image of the target.

Detection of Moving Targets in SAR Images. Various processing methods have been developed to detect and reposition moving targets:

Single-Aperture Moving-Target Indication (MTI) SAR With respect to conventional single-aperture SAR, key results have been obtained by several authors, including Raney,[38] Freeman,[39] Freeman and Currie,[40] and Werness et al.[41] If the PRF is greater than the minimum necessary to produce a SAR image, then further doppler bands are available. These bands can be used for additional information, and processing results for moving targets will be different from those for fixed targets. Freeman[39] presents a summary of potential results for moving targets, covering such issues as azimuth shift, range walk, and azimuth defocus: "Probably the worst defect...will be displacement of moving targets in the azimuth direction, away from their true position on the ground. The prefilter we have described is optimised for targets traveling radially...Such targets will appear at their correct position in the MTI image."

Interferometric SAR (InSAR) for Moving Target Indication (MTI) As mentioned in Section 17.3, Interferometric SAR (InSAR, sometimes also called IFSAR) refers to the use of two antennas whose signals are combined coherently. The two antennas are displaced horizontally (along a line parallel to the ground) to detect and analyze moving targets and are displaced vertically to estimate terrain height. Both types of InSAR are discussed herein. The former is discussed in this section and the latter is discussed later in "Interferometric SAR (InSAR) for Target Height Measurement."

InSAR to detect moving targets was originally developed by the Jet Propulsion Laboratory (JPL) to detect ocean currents,[18,19] and has been improved by several authors.[42] One of the most sophisticated techniques has been developed for the Joint STARS aircraft and uses an interesting combination of SAR and MTI techniques to detect and evaluate moving targets.[21]

The Joint STARS SAR mode involves a classical single-receiver-channel spotlight SAR that dwells on a designated ground-referenced coordinate for a duration that results in a nominally square point-spread function (i.e., downrange resolution = crossrange resolution). The MTI mode is capable of detecting and accurately geolocating both exoclutter (i.e., target moving faster than apparent terrain motion) and endoclutter (i.e., target moving more slowly than apparent terrain motion) returns from moving targets that, in general, have radar cross sections that are smaller than

the corresponding main-beam clutter-only cells. It accomplishes this by transmitting a coherent burst of pulses that are subsequently received by each of three linearly displaced subarrays (or interferometer ports). In each channel, the pulses are fast-time and slow-time processed into a set of range and doppler cells whose intensities may be considered as a SAR image of the scene (although not as fine in range or doppler resolution as in the SAR mode and generally not with a nominally square PSF). Each of the interferometer ports produces a (complex-valued) range-doppler "image" that could be called a "SAR" image, since it was formed from a coherent sequence of pulses, and the subsequent complex pair-wise combining of these images with the proper relative complex weighting to null the clutter can be considered as an InSAR process. On the other hand, to avoid confusion with the type of interferometric SAR that is used for target height measurement, the Joint STARS team typically refers to their process as "Clutter Suppression Interferometry" or simply "CSI".‡

Barbarossa and Farina[43] show that, by using multiple subapertures, detection and repositioning of moving targets can be considerably improved, in an extension of the real-beam Displaced Phase-Center Antenna (DPCA) technique (Staudaher[44]). They developed a procedure for SAR processing using an arbitrary number of subapertures, separated horizontally, to cancel ground clutter and image a moving target. Their approach is a combination of space-time processing (see the literature[1,25,45,46,] for a discussion of Space-Time Adaptive Processing or STAP) and time-frequency processing.[47,48] Figure 17.10*a* shows a simulated point-spread function (PSF) of a moving point target after clutter cancellation, and Figure 10*b* shows the final simulated PSF after range migration compensation.

The authors state: "The [point] target is supposed moving on the terrain (shadowing effects have been neglected) at a constant velocity, in a direction oblique with respect to the radar motion. The velocity parameters have been chosen in order to make evident the presence of range migration and of cross-range smearing of the target image. The ground reflectivity has been assumed equal to the target reflectivity (this is quite a pessimistic assumption, because in many cases of practical interest, the target reflectivity is higher). A receiver thermal noise, 40 dB below the target return, has also been summed to the received signal. The ground echo is first canceled, by using a two-element antenna and two time samples…The two antennas are separated by $d = vT$. [v = platform speed, T = pulse repetition interval.] A SAR image is then formed by conventional techniques. The result is shown in Figure [17.10*a*]. The smearing of the moving pointlike target is evident. Given the motion parameters, the target has migrated over six range cells. This is the cause for the broadening of the target image even in range, as well as in cross range.…The target echo causes a detection and initializes the motion estimation channel. The high [fine] resolution data are initially smoothed in range to decrease [coarsen] the range resolution. Then the processor looks for the range cell with the maximum energy content and computes the WVD [Wigner-Ville Distribution[48]] of that cell only.…The phase history is then used for compensating the range migration and the phase shift on the high [fine] resolution range data. The final image is shown in Figure [17.10*b*]. The sharpening of the target image is quite evident." Barbarossa and Farina assumed a point target in their simulation. They obtain a precise location of a simulated moving point target, but do not claim to have produced a simulated image of an extended moving target, such as a vehicle.

‡ Paragraph courtesy of Dr. Marshall Greenspan, Northrop-Grumman Corporation.

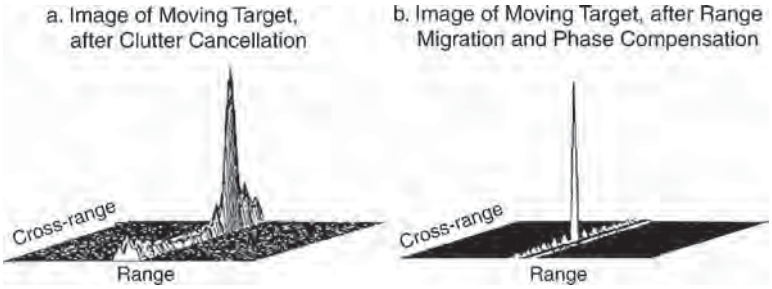


FIGURE 17.10 SAR image of a moving target using multiple subapertures for clutter cancellation (simulated data): (*a*) adaptive processing and (*b*) adaptive processing plus SAR (*after S. Barbarossa and A. Farina*[43] *© IEEE 1994*)

Guarino and Ibsen[49] describe an experiment using the AN/APG-76 radar. "The radar provides a unique simultaneous SAR/GMTI [Ground Moving Target Indication—scanning real aperture radar] mode in which detected targets are displayed on the SAR map as moving target symbols. The symbols are accurately located at their true azimuth position relative to the map center...The [SAR] map on which the moving target symbol is displayed is collected and processed simultaneously with the GMTI." The authors also make clear that Global Positioning System (GPS) inputs were essential to their accurate location of targets; fixed targets were geolocated to an absolute accuracy of better than 3 meters, and moving target vehicles to an absolute accuracy of about 15 meters. Stimson[5] also discusses these AN/APG-76 results (pp. 434, 554).

Imaging of Moving Targets in SAR Images. Perry et al.[50] have developed a method for SAR imaging of ground-moving targets that have unknown straight-line, constant-speed motion. They process the received phase history with a "keystone formatting" procedure that eliminates the effects of linear range migration for all ground-moving targets, regardless of their unknown velocity. The processing procedure then automatically focuses the moving targets. Figure 17.11*a* shows a conventionally processed SAR image containing three moving targets—a military truck (type M813), a tractor-trailer truck, and a surrogate (i.e., a full-size replica) of a missile transporter-erector-launcher (TEL). Figure 17.11*b* shows the focused image of the tractor-trailer resulting from the processing. The two-foot resolution clearly shows the outline of the cab and trailer of the truck.

Vibrating Targets in a SAR Image. Consider a SAR observing a scene that contains a point target whose position is oscillating sinusoidally (vibrating) (see Section 9.4 of Carrara et al.[3]). The component of the vibration amplitude that is parallel to the line-of-sight (LOS) is d. The variable component of the radar-target distance is then

$$\Delta R_{\text{tgt}}(t) = d \sin (2\pi f_{\text{vib}} t) \tag{17.41}$$

where f_{vib} is the vibration frequency.

Let the normalized complex echo corresponding to a stationary pixel be $e^{j2\pi f_d t}$, where f_d is the doppler frequency characteristic of the pixel. (We assume low fractional bandwidth, in which case this is equivalent to the doppler frequency corresponding to the center frequency of the transmitted bandwidth; see Section 8.1.5 of Sullivan[1]).

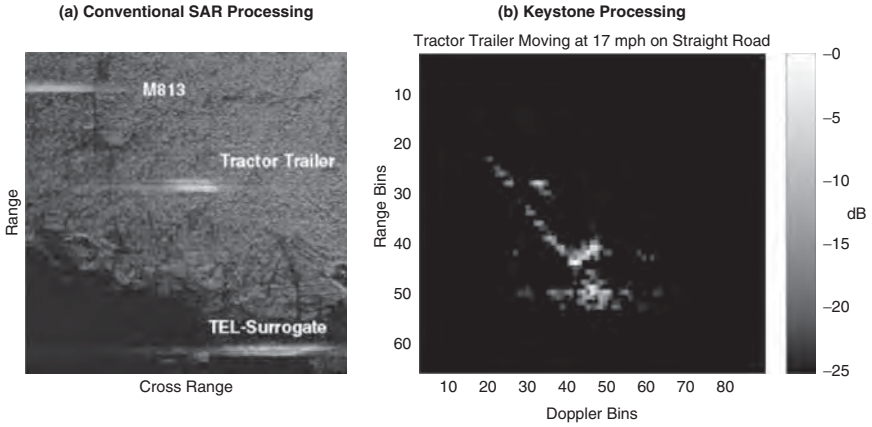


FIGURE 17.11 Imaging a moving target in a SAR image: (*a*) Conventional SAR image (real data) showing blurred moving vehicles: an M813 military truck, a tractor-trailer truck, and a surrogate (mockup) of a missile transporter-erector-launcher. (*b*) Image of tractor-trailer truck using "keystone" processing; the truck's cab is at the bottom and the trailer is above it. (*after R. P. Perry et al.*[50] *© IEEE 1999*)

In addition to this echo, the pixel containing the target will produce an additional echo with a periodic *phase error:*

$$\phi_e = \frac{4\pi d}{\lambda}\sin(2\pi f_{\text{vib}}t) = \phi_o \sin(2\pi f_{\text{vib}}t) \tag{17.42}$$

We assume $4\pi d \ll \lambda$, thus $\phi_o \ll 1$. The normalized complex echo corresponding to the phase error is

$$e^{j\phi_e} = e^{j\phi_0 \sin(2\pi f_{\text{vib}}t)} \tag{17.43}$$

$$\approx 1 + j\phi_0 \sin(2\pi f_{\text{vib}}t) \tag{17.44}$$

$$= 1 + \frac{\phi_0}{2}(e^{j2\pi f_{\text{vib}}t} - e^{-j2\pi f_{\text{vib}}t}) \tag{17.45}$$

The normalized complex echo corresponding to the vibrating target is then the complex product of the echo corresponding to the stationary pixel and the echo corresponding to the phase error:

$$e^{j\phi_{\text{dop}}} = e^{j2\pi f_d t} e^{j\phi_e} \tag{17.46}$$

and

$$e^{j\phi_{\text{dop}}} = e^{j2\pi f_d t} + \frac{\phi_o}{2}(e^{j2\pi t(f_d + f_{\text{vib}})} - e^{j2\pi t(f_d - f_{\text{vib}})}) \tag{17.47}$$

In the SAR image, the vibrating point target will appear in three crossrange locations. Most of the target energy still appears at the correct location, whereas a small fraction of the energy will appear in each of two pixels separated in crossrange by f_{vib} in doppler frequency. Thus, a vibrating target can give rise to a pair of distinctive echoes.

The corresponding velocity separation is $\Delta v = \pm f_{\text{vib}}\,\lambda/2$, and the crossrange displacement is then

$$\Delta r = \frac{\Delta v}{\Omega} = \frac{\Delta v}{V} R = \pm \frac{f_{\text{vib}} \lambda_{\text{avg}} R}{2\text{V}} \tag{17.48}$$

where λ_{avg} = average wavelength (assuming low fractional bandwidth). The relative amplitude of each of the paired echoes is

$$\text{Voltage:} \qquad \frac{\phi_o}{2} = \frac{2\pi d}{\lambda_{\text{avg}}} \tag{17.49}$$

$$\text{Power:} \qquad \left(\frac{\phi_o}{2}\right)^2 = \left(\frac{2\pi d}{\lambda_{\text{avg}}}\right)^2 \tag{17.50}$$

Thus, the amplitude of the paired echoes is proportional to the square of the vibration amplitude, and crossrange displacement is proportional to the vibration frequency (see Section 7.5.2 of Sullivan[1]).

For bright, point-like targets, or for targets such that $4\pi d$ is not small compared to λ, additional terms should be retained in Eq. 17.44, describing the fact that a series of paired echoes, of decreasing amplitude, may appear in crossrange.

Figure 17.12 shows a SAR image of a scene including a vibrating target—a truck with its engine running. The image contains two sets of paired echoes, corresponding to two vibration frequencies characteristic of the particular truck used for the observation.

Measurement of Object Height. The basic theory of SAR assumes that the scene is flat. To the extent that the scene is not flat, distortions in the SAR image will result. In some cases, they can be used to measure the height of elevated objects above a flat terrain.

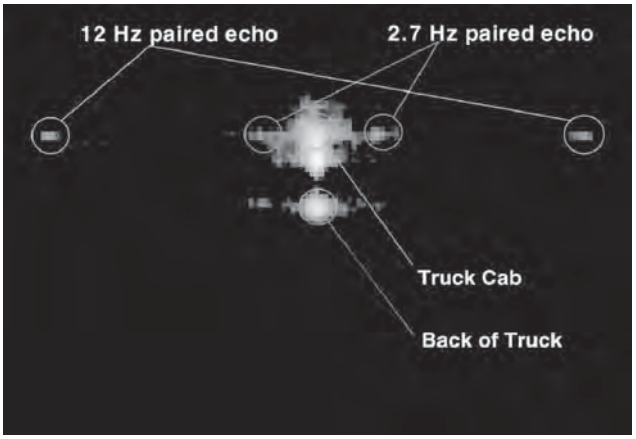


FIGURE 17.12 SAR scene containing vibrating target: Crossrange is horizontal; range is vertical. The paired echoes in crossrange are characteristic of a vibrating target (truck with engine running) in a SAR image. The frequencies (2.7 Hz and 12 Hz) are peculiar to the specific truck that was imaged. (*Courtesy of Northrop-Grumman Corporation*)

Shadows. The simplest method of measuring object height is to observe the length L_{shadow} of the shadow of the object cast by the SAR and calculate the object height h from the known SAR altitude H and ground range R_g:

$$h = L_{\text{shadow}} \cdot \frac{H}{R_g} \tag{17.51}$$

This expression assumes flat-earth, and may be generalized to curved-earth if R_g is relatively large; see Section 3.2.2 of Sullivan.[1] However, the shadow method works only for an isolated, relatively high object on essentially flat terrain (e.g., Figure 17.9).

Layover. SAR processing sorts target returns into bins (pixels) depending on the range R and velocity v of the target relative to the platform. If two or more targets have the same R and v, then they will be placed at the same location in the SAR image.

We shall define a *layover contour* as the locus of points in 3D space such that an object at any of the points will be assigned to the same location in a SAR image. As shown in Figure 17.13, a layover contour is the intersection of a constant-range sphere of range R and a constant velocity cone of generating angle $\beta = \cos^{-1}(v/V)$ with axis along the platform direction, i.e., a circle of radius $R \sin\beta$ ahead of the platform. ($\beta > 90°$ corresponds to targets behind the platform.) We shall, therefore, call the contour the *layover circle*. If the top of an elevated object, such as a tower, is on the layover circle and if the ground is flat, then the top of the tower will appear in the SAR image at the same position as a point on the ground where the layover circle intersects the ground. The tower will be "laid over," hence the nomenclature.

As shown in Figure 17.13, let us consider a platform in straight-line, constant-speed motion at altitude H over a flat earth forming a SAR image, the center of which is at slant range $R_s \gg H$ and squint angle θ_{sq}. Suppose that a tower of height h ($h \ll H$) is in the area that is imaged. We describe locations within the image by a coordinate system (x_1, y_1). If the base of the tower is at (x_{10}, y_{10}), we wish to ascertain the image coordinates of the top of the tower.

Figure 17.13*a* illustrates a perspective view. Because $R_g \gg H$, the isodop (y_1 – axis) makes an angle θ_{sq} with the y-axis. The image center is a distance $S \cong R_g \cos\theta_{\text{sq}}$ from the x-axis, where R_g = ground range. Figure 17.13*b* shows a view from the $+x$-axis, indicating the layover circle normal to the x-axis and showing that the image location of the tower top is located a distance (the *layover distance*) $d \cong hH/(R_g \cos\theta_{\text{sq}})$ from the image location of the tower base. Figure 17.13*c* then depicts the view in image coordinates (x_1, y_1). The image coordinates of the tower top are

$$x_{11} = x_{10} + d \sin\theta_{\text{sq}},\ y_{11} = y_{10} - d \cos\theta_{\text{sq}}$$

For example, if R_g = 100 km, H = 5 km, h = 100 m, and $\theta_{\text{sq}} = 0$, then d = 5 m, $x_{11} = x_{10}$, and $y_{11} = y_{10} - 5$ m. The tower top appears in the image 5 m closer to the radar than the tower base. This principle may sometimes be used to estimate the height of isolated, tower-like structures on relatively level ground:

$$h \cong \frac{dR_g \cos\theta_{\text{sq}}}{H} \tag{17.52}$$

The intersection of the constant-velocity cone and the ground is a hyperbola. If H is not $\ll R$, the isodop direction will not be parallel to the down-range direction.

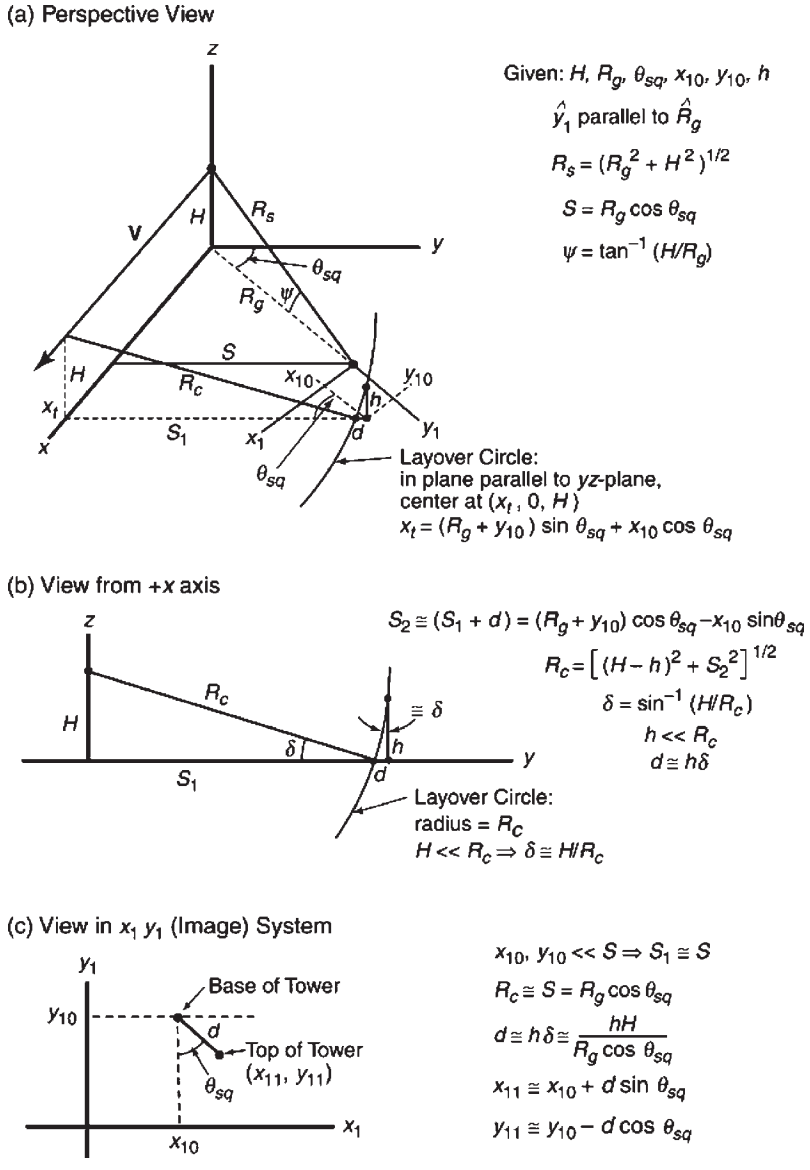


FIGURE 17.13 Layover: (*a*) Perspective view, (*b*) view along platform flight path, and (*c*) view in coordinate system of SAR image (*Courtesy of SciTech Publishing, Inc.*)

The geometry is more complicated but layover distance and object height may still be estimated.

Stereo SAR. Two SAR images of the same scene may be obtained from somewhat different locations (see Section 9.3.8.1 in Carrara et al.[3]). Noncoherent comparison

of the two—the *stereo* technique—may enable estimation of object height. The technique is analogous to the method by which we humans use our two eyes to help estimate the distance of the objects that we see. In fact, the two SAR images may be printed on the same page using two different colors, with the viewer using special glasses so that the left eye sees only one image and the right eye only the other and the brain processing the two together so that the scene is perceived in 3D.

Interferometric SAR (InSAR) for Target Height Measurement. Interferometric SAR (InSAR)[42,51,52] (also called IFSAR; see Section 9.3 of Carrara et al.[3] and Adams et al.[20]), when used for terrain height measurement, involves two SAR images taken from antennas at slightly different altitudes and compared *coherently* to obtain fine-resolution information regarding the height of terrain or targets in the image. (In this case, it is sometimes called IFSARE, where *E* emphasizes elevation measurement.) InSAR may be performed using a single platform with two antennas (*single-pass InSAR*) or by the same platform making two passes over the same terrain (*two-pass InSAR*). Allen[42] gives several examples of fielded systems using each type. It is essential that the relative locations of the two antennas be rather precisely known. Advantages and disadvantages of the two types of InSAR are as follows.

Two-pass InSAR:

- No special hardware is required; a conventional SAR may be flown twice over the designated terrain.
- Motion compensation is challenging; the position of the antenna versus time in each pass must be known with great precision.
- A long baseline (vertical distance between antenna paths) provides fine vertical resolution (but challenging ambiguities).
- The scene may change between passes due to wind, etc.
- Example results are given by Schuler et al.,[53] who perform "terrain topography measurement using multipass polarimetric SAR."

Single-pass InSAR:

- Baseline is relatively well known, providing consistency throughout synthetic aperture.
- Scene is same for both images since data for each are collected simultaneously.
- On-board, real-time processing is a possibility.
- More sophisticated (expensive) hardware is required: two antennas, two receiver channels, and two sets of analog-to-digital (A/D) converters.
- Example results are given by Adams et al.,[20] who include an InSAR image of the stadium at the University of Michigan in Ann Arbor, viewable with two-color glasses.

To understand the theory of InSAR, we first consider two antennas, *A* and *B*, separated vertically by a baseline *L*, observing a point target *a* that is on a flat ground at range *R*; the radar line-of-sight (LOS) intersects the flat ground at grazing angle ψ_1 (Figure 17.14). We consider two possibilities: (1) one antenna transmits and each receives ($n = 1$), and (2) antenna *A* transmits and then receives; then antenna *B* transmits and then receives ($n = 2$).

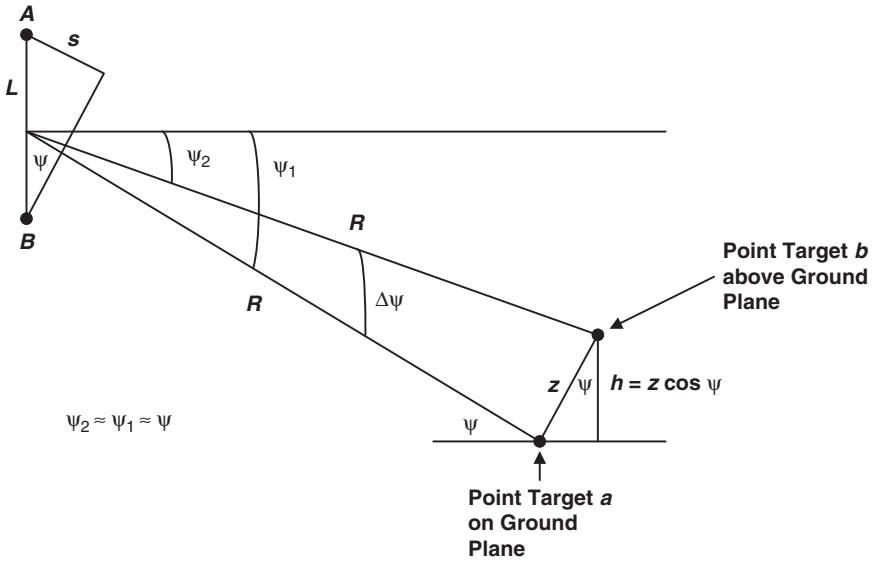


FIGURE 17.14 InSAR—-vertical antenna separation: Vertical antenna separation enables estimation of the height of a target above the average ground height or, more generally, terrain height versus location. (*Courtesy of SciTech Publishing, Inc.*)

For a "single-frequency" pulse of wavelength λ, the difference in the phases of the echoes received from the point target observed by the two antennas is (from Figure 17.14)

$$\Delta\phi_1 = \frac{2\pi ns}{\lambda} = \frac{2\pi nL\sin\psi_1}{\lambda} \tag{17.53}$$

We now consider the same antennas A and B observing a second point target b that is a distance h above the flat ground also at range R; the radar line-of-sight (LOS) intersects the parallel to the flat ground at grazing angle ψ_2. The difference in the phases of the echoes received from the point target observed by the two antennas is now

$$\Delta\phi_2 = \frac{2\pi nL\sin\psi_2}{\lambda} \tag{17.54}$$

We consider the quantity

$$\Delta\phi \equiv |\Delta\phi_2 - \Delta\phi_1| = \frac{2\pi nL}{\lambda}|\sin\psi_2 - \sin\psi_1| \tag{17.55}$$

We assume that $h \ll R$ and thus $\psi_2 \approx \psi_1 \approx (\psi_2 + \psi_1)/2 = \psi$. Then, where $\Delta\psi = |\psi_2 - \psi_1|$, from Figure 17.14,

$$\Delta\phi \approx \frac{2\pi nL}{\lambda}\cos\psi\Delta\psi = \frac{2\pi nL}{\lambda}\cos\psi\frac{h}{R\cos\psi} = \frac{2\pi nLh}{\lambda R} \tag{17.56}$$

We now consider $|\delta(\Delta\phi)|$ = the change in $\Delta\phi$ due to a change in h given by $|\delta h|$:

$$|\delta(\Delta\phi)| = \frac{2\pi nL}{\lambda R}|\delta h| \tag{17.57}$$

Antennas A and B may be considered as separated in vertical distance on an aircraft. The relationship between $\delta(\Delta\phi)$ and a variation in terrain altitude δh is

$$|\delta h| = \frac{\lambda R\,|\delta(\Delta\phi)|}{2\pi nL} \qquad \text{(antennas separated vertically)} \tag{17.58}$$

InSAR for terrain elevation measurement can also be performed by an aircraft with two antennas separated *horizontally* (perpendicular to the flight path) by L. The aircraft is banking at angle $\gamma > 0$ and collecting data from the flat ground at grazing angle ψ (Figure 17.15). Then the effective aperture (perpendicular to the LOS) is $L\sin(\psi+\gamma)$ instead of $L\cos\psi$. From Figure 17.15, we have

$$|\delta h| = \frac{\lambda R\,|\delta(\Delta\phi)|\cos\psi}{2\pi nL\sin(\psi+\gamma)} \qquad \text{(antennas separated horizontally)} \tag{17.59}$$

In either geometry, since both channels are noisy, the expected accuracy in the phase difference (1-sigma) is given by Levanon[17]:

$$\delta(\Delta\phi) = \frac{1}{\sqrt{\text{SNR}}}\sqrt{2} \tag{17.60}$$

where SNR = signal-to-noise ratio (see Section 17.6). Thus, the theoretical accuracy for terrain altitude measurement is

Vertical antenna separation (no banking):

$$\delta h = \frac{\lambda R}{\pi nL\sqrt{2\text{SNR}}} \tag{17.61}$$

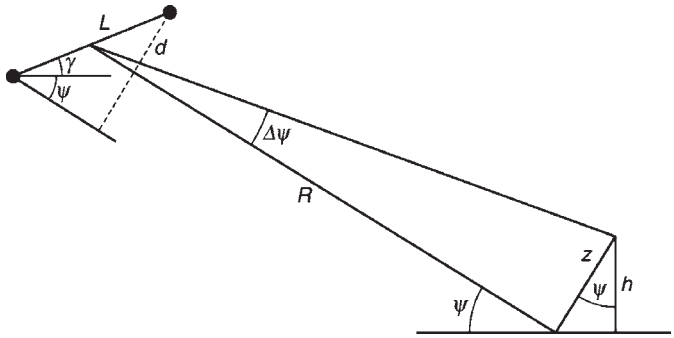


FIGURE 17.15 InSAR—horizontal antenna separation: As long as the two apertures do not both lie on the same line of sight to the target area, an estimate of terrain height may be made, by comparing the phases of echoes received by the two apertures. (*Courtesy of SciTech Publishing, Inc.*)

Horizontal antenna separation:

$$\delta h = \frac{\lambda R \cos\psi}{\pi n L \sin(\psi + \gamma)\sqrt{2\text{SNR}}} \tag{17.62}$$

Furthermore, when the phase moves through an interval of 2π, an ambiguity occurs in terrain altitude measurement. The corresponding altitude difference is computed by replacing $\delta(\Delta\phi)$ by 2π in Eqs. 17.58 and 17.59.

Vertical antenna separation (no banking):

$$\Delta h(ambig) = \frac{\lambda R}{nL} \tag{17.63}$$

Horizontal antenna separation:

$$\Delta h(ambig) = \frac{\lambda R \cos\psi}{nL \sin(\psi + \gamma)} \tag{17.64}$$

Although we have derived these relationships for a single monochromatic pulse, they can be shown (see Section 8.1.5 of Sullivan[1] and Sections 3.0 and 9.3 of Carrara et al.[3]) to be true for SAR pixels also, with λ replaced by c/f_{avg}.

The National Aeronautics and Space Administration (NASA) performed successful X/C-band single-pass polarimetric InSAR (single-pass) for the Space Shuttle Radar Topography Mission (SRTM), using antennas on the shuttle itself and on the shuttle's maneuverable arm to produce a complete 3D map of the Earth's surface between 60° N latitude and 56° S latitude—nearly 80% of the surface, with best vertical accuracy of 6 meters on a 30-meter horizontal grid.[54]

Foliage-Penetration (FOPEN) SAR. Although higher frequency (greater than approximately 2 GHz) microwaves do not penetrate foliage well, lower frequency microwaves do (Fleischman et al.[55]; see also Section 21-6 of Ulaby et al.[56]). For example, for C band, the attenuation of a typical forest canopy varies from ≈10 dB to ≈40 dB; the probability that the attenuation is less than 20 dB is about 0.2. On the other hand, for UHF radiation, attenuation varies from 0 to ≈20 dB; half the time it is less than ≈7 dB. Thus, for FOPEN, UHF radiation is necessary: shorter waves will not penetrate the foliage, whereas for airborne applications, longer waves would require prohibitively large antennas. (Specific values of attenuation [dB/meter] vary with grazing angle, tree type, leaf density, and moisture content; however, the previous statement is a general summary of these results; further details are given in Fleishman et al.[55])

The aperture time t_A required to collect sufficient data for a SAR image is found from Eq. 17.6. For example, let us assume $R = 100$ km, $V = 180$ m/sec (350 kts), $\theta_{\text{sq}} = 0$, and $\delta_{\text{cr}} = 1$ m. For $f = 10$ GHz (X band, $\lambda = 0.03$ m), $t_A = 8.3$ sec, and the fractional bandwidth $B/f_o = 0.015$. On the other hand, for $f_o = 0.5$ GHz (UHF, $\lambda = 0.6$ m), $t_A = 167$ sec = 2.8 min and $B/f_o = 0.3$. Such a high fractional bandwidth (an *ultra-wideband SAR*) presents challenges in designing hardware components, such as antennas, that are reasonably linear over the full frequency range.[57] Furthermore, the long aperture time presents motion-compensation challenges, and the wide real-beam angle adds to processing difficulties, very likely requiring Range-Migration Algorithm (RMA) processing (Section 17.3). In addition, for calculating crossrange resolution, the small-angle approximation no longer holds.

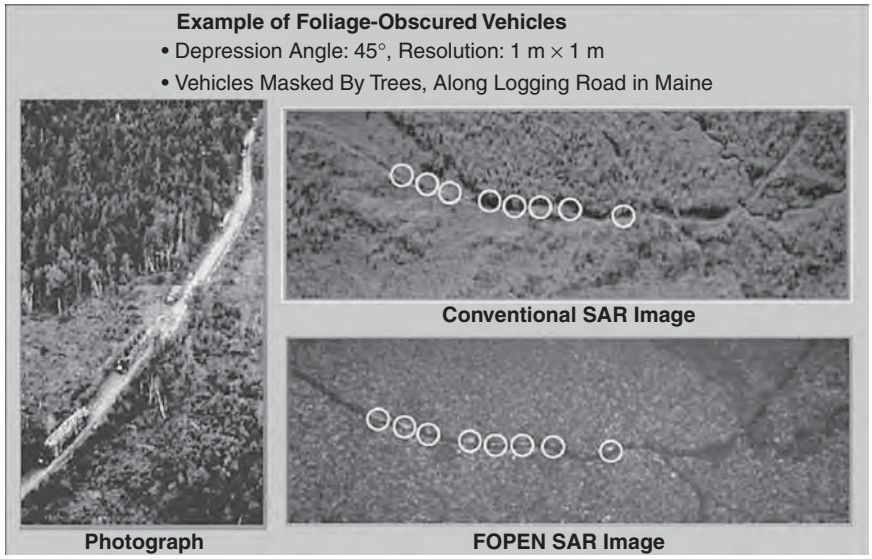

Example of Foliage-Obscured Vehicles
• Depression Angle: 45°, Resolution: 1 m × 1 m
• Vehicles Masked By Trees, Along Logging Road in Maine
Conventional SAR Image
Photograph
FOPEN SAR Image

FIGURE 17.16 Conventional and FOPEN SAR[62] images: In the conventional image, target vehicles cannot readily be seen, whereas in the FOPEN image they are quite prominent. (*Source: http://www.darpa .mil/DARPATech2000/Presentations/spo_pdf/4MoyerCCTB&W.pdf*)

A number of authors report successful results with FOPEN SAR, using such SARs as the Environmental Research Institute of Michigan P-3 SAR,[58,59] the Swedish National Defense Research Establishment CARABAS sensor,[60] and the SRI International Ultra-Wideband SAR.[61] Furthermore, Moyer[62] presents images from conventional and FOPEN SARs showing that vehicles under trees may be imaged significantly better with FOPEN SAR than with conventional SAR; example imagery is given in Figure 17.16.

REFERENCES

1. R. J. Sullivan, *Radar Foundations for Imaging and Advanced Concepts*, Raleigh, NC: SciTech, 2004; previously published as *Microwave Radar: Imaging and Advanced Concepts*, Norwood, MA: Artech House, 2000.
2. L. J. Cutrona, "Synthetic aperture radar," in M. Skolnik, *Radar Handbook*, 2nd Ed., New York: McGraw-Hill, 1990; 1st Ed., New York: McGraw-Hill, 1970.
3. W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar*, Norwood, MA: Artech House, 1995.
4. J. Curlander and R. McDonough, *Synthetic Aperture Radar*, New York: John Wiley and Sons, 1991.
5. G. W. Stimson, *Introduction to Airborne Radar*, 2nd Ed., Mendham, NJ: SciTech, 1998.
6. C. J. Jakowatz, Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode SAR: A Signal-Processing Approach*, Boston: Kluwer Academic Publishers, 1996.
7. S. A. Hovanessian, *Introduction to Synthetic Array and Imaging Radars*, Norwood, MA: Artech House, 1980.

8. R. O. Harger, *Synthetic Aperture Radar Systems: Theory and Design*, New York: Academic Press, 1970.
9. R. Birk, W. Camus, E. Valenti, and W. McCandless, "Synthetic aperture radar imaging systems," *IEEE AES Magazine*, pp. 15–23, November 1995.
10. C. Jackson and J. Apel (deceased; the book is dedicated to him), *Synthetic Aperture Radar Marine User's Manual*, Washington, DC: Department of Commerce, National Oceanic and Atmospheric Administration (NOAA), 2004.
11. D. Ausherman, A. Kozma, J. Walker, H. Jones, and E. Poggio, "Developments in radar imaging," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-20, no. 4, July 1984.
12. C. Wiley, "Synthetic aperture radars," *IEEE Transactions Aerospace and Electronic Systems*, vol. AES-21, pp. 440–443, May 1985.
13. L. J. Cutrona, W. E. Vivian, E. N. Leith, and G. O. Hall, "A high-resolution radar combat-surveillance system," *IRE Transactions on Military Electronics*, vol. MIL-5, no. 2, pp. 127–131, April 1961. (Reprinted in Kovaly.[15])
14. C. W. Sherwin, J. P. Ruina, and R. D. Rawliffe, "Some early developments in synthetic aperture radar systems, *IRE Transactions on Military Electronics*, vol. MIL-6, no. 2, pp. 111–115, April 1962. (Reprinted in Kovaly.[15])
15. J. J. Kovaly, *Synthetic Aperture Radar*, Norwood, MA: Artech House, 1976. (This is a collection of early classic papers concerning SAR.)
16. D. C. Schleher, *MTI and Pulsed Doppler Radar*, Norwood, MA: Artech House, 1991.
17. N. Levanon, *Radar Principles*, New York: Wiley-Interscience, 1988.
18. R. M. Goldstein and H. A. Zebker, "Interferometric radar measurement of ocean currents," *Nature*, vol. 328, pp. 707–709, 1987.
19. R. M. Goldstein, H. A. Zebker, and T. P. Barnett, "Remote sensing of ocean currents," *Science*, vol. 246, pp. 1282–1285, 1989.
20. G. F. Adams et al., "The ERIM interferometric SAR: IFSAR," in *Proceedings of the 1996 IEEE National Radar Conference*, 1996, pp. 249–254. (Reprinted in *IEEE AES Systems Magazine*, December 1996.)
21. M. Skolnik, *Introduction to Radar Systems*: 1st Ed., New York: McGraw-Hill, 1962; 2nd Ed., New York: McGraw-Hill, 1980; 3rd Ed., New York: McGraw-Hill, 2001.
22. S. Musman, D. Kerr, and C. Bachmann, "Automatic recognition of ISAR ship images," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1392–1404, October 1996.
23. W. L. Wolfe and G. Zissis (eds.), *The Infrared Handbook*, rev. ed., Ann Arbor, MI: Environmental Research Institute of Michigan (now General Dynamics, Ypsilanti, MI), 1989.
24. M. Richards, *Fundamentals of Radar Signal Processing*, New York: McGraw-Hill, 2005.
25. R. Klemm, *Principles of Space-Time Adaptive Processing*," London: IEE, 2002.
26. E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, 2nd Ed., Raleigh, NC: SciTech, 2004.
27. E. O. Brigham, *The Fast Fourier Transform and Its Applications*, Englewood Cliffs, NJ: Prentice Hall, 1988.
28. E. D. Kaplan, *Understanding GPS, Principles and Applications*, Norwood, MA: Artech House, 1996.
29. R. W. Bayma and P. A. McInnes, "Aperture size and ambiguity constraints for a synthetic aperture radar," in *Proc. 1975 International Radar Conference*, pp. 499–504. (Reprinted in Kovaly.[15])
30. F. M. Henderson and A. J. Lewis (eds.), *Principles and Applications of Imaging Radar*, New York: Wiley, 1998.
31. C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Norwood, MA: Artech House, 1998.
32. F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, January 1978.
33. D. Barton, *Radar Systems Analysis and Modeling*, Norwood, MA: Artech House, 2004.

34. R. J. Sullivan, A. D. Nichols, R. F. Rawson, C. W. Haney, F. P. Dareff, and J. J. Schanne, Jr., "Polarimetric X/L/C-band SAR," in *Proceedings of the 1988 IEEE National Radar Conference*, 1988, pp. 9–14.
35. D. N. Held, W. E. Brown, and T. W. Miller, "Preliminary results from the NASA/JPL multifrequency, multipolarization SAR," in *Proceedings of the 1988 IEEE National Radar Conference*, 1988, pp. 7–8. *See also* P. A. Rosen et al., "UAVSAR: New NASA airborne SAR system for research," *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 11, pp. 21–28, November 2007.
36. L. M. Novak, M. C. Burl, and W. W. Irving, "Optimal polarimetric processing for enhanced target detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 234–243, January 1993.
37. L. M. Novak, S. D. Halversen, G. J. Owirka, and M. Hiett, "Effects of polarization and resolution on SAR ATR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 1, pp. 102–115, January 1997.
38. R. K. Raney, "Synthetic aperture imaging radar and moving targets," *IEEE Transactions of Aerospace and Electronic Systems*, vol. AES-7, no. 3, pp. 499-505, 1971.
39. A. Freeman, "Simple MTI using synthetic aperture radar," in *Proceedings of IGARSS 1984 Symposium*, ESA SP-215, 1984.
40. A. Freeman and A. Currie, "Synthetic aperture radar (SAR) images of moving targets," *GEC J*. Res., vol. 5, no. 2, pp. 106–115, 1987.
41. S. Werness, W. Carrara, L. Joyce, and D. Franczak, "Moving target algorithms for SAR data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-26, no. 1, pp. 57–67, 1990.
42. C. T. Allen, "Interferometric synthetic aperture radar," *IEEE GRS Society Newsletter*, pp. 6–13, November 1995.
43. S. Barbarossa and A. Farina, "Space-time-frequency processing of synthetic aperture radar signals," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 2, pp. 341–358, April 1994.
44. F. M. Staudaher, "Airborne MTI," Chapter 16 in *Radar Handbook*, M. Skolnik (ed.), 2nd Ed., New York: McGraw-Hill, 1990.
45. J. Ward, *Space-Time Adaptive Processing for Airborne Radar*, Technical Report 1015, Lexington, MA: Lincoln Laboratory, Massachusetts Institute of Technology, 1994.
46. J. Guerci, *Space-Time Adaptive Processing for Radar*, Norwood, MA: Artech House, 2003.
47. V. C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*, Norwood, MA: Artech House, 2002.
48. L. Cohen, "Time-frequency distributions–a review," *Proceedings of the IEEE*, vol. 77, no. 7, July 1989.
49. R. Guarino and P. Ibsen, "Integrated GPS/INS/SAR/GMTI radar precision targeting flight test results," in *Proceedings Institute of Navigation GPS-95 Conference*, 1995, pp. 373–379.
50. R. P. Perry, R. C. DiPietro, and R. L. Fante, "SAR imaging of moving targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 188–200, January 1999.
51. J. Rodriguez and J. M. Martin, "Theory and design of interferometric synthetic aperture radar," *IEE Proceedings*, Part F, vol. 139, pp. 147–159, April 1992.
52. R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, pp. R1 to R54, August 1998. *See also* F. Gini and F. Lombardini, "Multibaseline cross-track SAR interferometry: A signal-processing perspective," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, Part 2: Tutorials, pp. 71–93, August 2005 ; M. A. Richards, "A beginner's guide to interferometric SAR concepts and signal processing," *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 9, Part 2: Tutorials, pp. 5–29, September 2007.
53. D. L. Schuler, J-S Lee, T. L. Ainsworth, and M. R. Grunes, "Terrain topography measurement using multipass synthetic aperture radar data," *Radio Science*, vol. 35, no. 3, May–June 2000, pp. 813–832.
54. W. B. Scott, "Flight to radar-map Earth from space," *Aviation Week and Space Technology*, pp. 50–53, September 20, 1999 (Cover Story).

55. J. G. Fleischman, S. Ayasli, E. M. Adams, D. R. Gosselin, M. F. Toups, and M. A. Worris, "Foliage penetration experiment," (series of three papers), *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 134–166, January 1996. (This series of papers was awarded the 1996 M. Barry Carlton Award; see *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, p. 1472, October 1999.)

56. F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave Remote Sensing*, 3 Volumes, Norwood, MA: Artech House, 1986.

57. E. L. Ayers, J. M. Ralston, R. P. Mahoney, P. G. Tomlinson, and J. McCorkle, "Antenna measures of merit for ultra-wide synthetic aperture radar," in *Proceedings of the 1998 IEEE Radar Conference*, 1998, pp. 331–336.

58. N. Vandenberg, D. R. Sheen, S. Shackman, and D. Wiseman, "P-3 ultrawideband SAR: System applications to foliage penetration," *Proceedings SPIE*, vol. 2757, pp. 130–135, 1996.

59. M. F. Toups, L. Bessette, and B. T. Binder, "Foliage penetration data collections and investigations utilizing the P-3 UWB SAR", *Proceedings SPIE*, vol. 2757, p. 136–144, 1999.

60. L. M. H. Ulander and P. O. Frolind, "Precision processing of CARABAS HF/VHF-band SAR data," *Proceedings IEEE Geoscience Remote Sensing Symposium IGARSS 1999*, Hamburg, Germany, vol. 1, 1999, p. 47–49. Also see L.M. Ulander et al., "Detection of concealed ground targets in CARABAS SAR images using change detection," *Proceedings SPIE*, vol. 3721, p. 243–252, *Algorithms for Synthetic Aperture Radar Imagery VI*, E. G. Zelnio (ed.), 1999.

61. E. M. Winter, M. J. Schlangen, and C. R. Hendrikson, "Comparisons of target detection in clutter using data from the 1993 FOPEN experiments," *Proceedings SPIE*, vol. 2230, p. 244–254, Algorithms for Synthetic Aperture Radar Imagery, D. A. Giglio (ed.), 1994.

62. L. Moyer, "Counter concealed target technologies," presented at DARPA Tech 2000, http://www.darpa.mil/DARPATech2000/Presentations/spo_pdf/4MoyerCCTB&W.pdf.

Chapter 18
Space-Based Remote Sensing Radars

R. Keith Raney

Johns Hopkins University Applied Physics Laboratory

18.1 PERSPECTIVE

Motivation. Worldwide, the rate of investment in space-based radar (SBR) during the time the radars in this chapter were operating was on the order of one billion dollars per year. Space-based synthetic aperture radar (SAR) systems capable of 1-m resolution have become the norm, with systems under development or already launched by at least seven countries. As expanded in the appropriate sections of this chapter, the (range) measurement precision of surface height change is now on the order of 1 millimeter per year, as established by two different classes of Earth-observing SBRs (SARs and radar altimeters). Several nations are sponsoring radars for exploration at the Moon and beyond. SBR-related patent applications continue apace. In short, SBR is an exciting, exacting, extensive, and expanding topic.

Space-based radar systems face fundamental challenges. The permissible options for the value of several parameters (such as pulse repetition frequency) are more restricted for SBR than for airborne systems. Likewise, the hardware environment imposes more rigorous constraints on implementation, and SBR systems do not enjoy the luxuries of hands-on maintenance nor in-field parts replacement. However, the payoff from radars in space more than compensates for these challenges, since space offers a unique perspective for Earth observation and is an essential viewpoint for lunar or planetary exploration.

Covered and Omitted Topics. This chapter introduces space-based remote sensing radars. The focus is on Type II SBRs, as outlined in the previous edition of this *Handbook*, including both Earth-orbiting and planetary systems. The material in this chapter is designed to be reasonably complete at a survey level; the discussion zooms in on selected case examples to illustrate application-specific implementations or technological innovations. An outstanding early example is the Seasat satellite launched in 1978 (Figure 18.1), which (as its name suggests) was designed for oceanic observations. Three of its sensors were SBRs—a synthetic aperture radar, an altimeter, and a scatterometer. As the reader will discover in this chapter, these three Seasat instruments established the initial paradigm for virtually all subsequent radars of their respective classes.[1,2]

FIGURE 18.1 The Seasat satellite, featuring the antennas of its three radars (*Courtesy of NASA*)

This chapter does not cover the Type I short-range specialized SBR systems reviewed in earlier editions, such as terminal guidance or rendezvous radars. Also not covered are Type III SBRs, such as multi-spacecraft systems for space-based radar surveillance of the Earth's surface or airspace. (Although these large SBR concepts are interesting in principle, their implied costs remain a disincentive, particularly if their erstwhile sponsors expect their performance to approach the current state-of-the-art in airborne search or surveillance radars.) Except for the mention of a few basic concepts, this chapter does not delve into the extensive subjects of orbitology, implementation of space-qualified hardware, nor system integration and test, for which the interested reader is directed to standard references.[3,4]

Basic Orbit Characteristics. Unlike an airborne platform that can go anywhere at any time (subject to fuel and air-space limitations), a satellite's position and velocity when in orbit about a planetary body are rigidly governed by orbital dynamics, summarized compactly by Kepler's laws. Further, access by an SBR to a given area of interest depends on the rotation rate of the planet, as well as the satellite's position along its orbit and the radar's viewing geometry. Thus, the primary parameters to be included in SBR mission design include orbital altitude, spacecraft on-orbit velocity (and hence period), orbit inclination, and the rotation rate of the planet.

Earth-observing SBRs, such as SARs, tend to operate from spacecraft in near-circular low-Earth orbits (LEO). Typical LEO altitudes span 500 km to 850 km. Lower altitudes incur larger atmospheric drag, whereas higher altitudes imply higher radiation levels and longer radar ranges, neither of which are desirable in most situations. Spacecraft velocities at these LEO altitudes are on the order of 7.5 km/s; their corresponding periods are about 100 minutes. Earth rotation (at a rate of ~0.25 degree/min) shifts the sub-satellite point at the equator by ~3300 km orbit-to-orbit. The orbit altitude normally may be chosen to tune the period and Earth rotation rate so that an exact-repeat pattern develops that has a stipulated number of days. The repeat period

often is an integral number, such as the 24-day repeat of RADARSAT, although a non-integral period may be preferred, such as the 9.916-day repeat of the TOPEX/Poseidon radar altimeter. Orbit parameters such as repeat period must be maintained, requiring small boosts from spacecraft thruster maneuvers,[4] typically applied every several weeks for LEO missions. The revisit time of a given SBR asset depends on the range swath width covered by the radar in question, as well as the orbit's exact-repeat period and the latitude of the site of interest. The off-nadir* look angle of certain SBRs can be adjusted to increase the effective revisit rate. Note that orbit maneuvers are not a practical means to improve the frequency of site coverage, since a large change in altitude (hence velocity), or especially any change of orbit inclination (the angle of the inertial orbit plane with respect to the Earth's equator), would require a substantial expenditure of precious onboard fuel resources.[3]

The Earth is flattened at the poles, due primarily to its relatively rapid rotation. The resulting lack of spherical symmetry in the gravity field at orbital altitude imposes small lateral forces on an inclined orbit plane, which consequently precesses in inertial space. The amount, direction, and rate of orbit precession can be controlled by choice of the inclination and the mean altitude of the orbit. Many satellite platforms use this degree of freedom to generate a sun-synchronous orbit, which is one that maintains a constant angle of its orbit plane relative to solar illumination over the entire year. The European Space Agency's Envisat spacecraft is a good example of a sun-synchronous LEO, having ~98.5° inclination† and 785 km altitude. Sun-synchronous spacecraft that host optical instruments (such as Japan's ALOS) choose the phase of the sun angle to favor illumination of the surface, which usually leads to a midday orbit from which most of the Earth's surface is viewed at about the same local time, near midday. Such orbits imply that the spacecraft must pass through the Earth's shadow about half the time, which has consequences on the design of the thermal and power subsystems in particular. In contrast, spacecraft that carry only radars, such as RADARSAT, tend toward favorable illumination of the spacecraft. The natural result in this case is the so-called dawn-dusk orbit, in which the satellite—and its solar-panel-dependent power system—avoids the shadow of the Earth almost always for almost all seasons.

Certain applications are particularly well-served by exact-repeat orbits. For example, if a group of orbital trajectories falls within a small radius of each other when over an area of interest, then radar measurements from several orbits may be compared coherently, thus potentially sensitive at the order of a wavelength to changes in the scene between observations. Such coherent change detection is a standard technique in the field of space-based SAR interferometry, reviewed in the following section. Exact-repeat orbits are standard for most radar altimeters, but for geophysical reasons, not for mutual coherency. Sun-synchronicity presents its own problems for ocean-sensing altimeters. These comments are elaborated in Section 18.3.

As discussed in several chapters of this book, the performance of doppler-sensitive radars is conditional upon the velocity of their host platform. The velocity of a spacecraft in orbit at altitude h above a planet of radius R_P and mass M_P is given by

$$V_{SC} = \sqrt{M_P G / (R_P + h)} \tag{18.1}$$

* *Nadir* is the point below the spacecraft on the surface intersected by the radius vector from the Earth's center to the spacecraft.

† Inclinations greater than 90° are retrograde because their E-W velocity component in the ascending pass is contrary to the direction of the Earth's rotation, in contrast to prograde orbits whose inclinations are less than 90°.

TABLE 18.1 Spacecraft Velocities

Body	Mass (kg)	Radius (km)	Altitude h (km)	V_{SC} (m/s)	hV_{SC} (km^2/s)
Earth	5.97×10^{24}	6380	800	7466	6000
Venus	4.87×10^{24}	6052	300	7151	2200
Mars	6.4×10^{23}	3397	400	3353	1600
Ganymede	1.4×10^{23}	2631	100	1849	185
Calisto	1.08×10^{23}	2400	100	1697	170
Moon	7.35×10^{22}	1737	100	1634	160
Europa	4.8×10^{22}	1569	100	1385	140

where G is the universal gravity constant[‡] 6.67×10^{-11} Nm^{2}kg^{-2}. Table 18.1 lists representative spacecraft velocities for bodies in the solar system that have been visited, or are likely to be observed, by range-doppler radars. Feasible satellite altitudes are limited below by the prevailing atmospheric density. The final column of the table lists the altitude-velocity product hV_{SC} corresponding to each entry. This product is a scaling factor that characterizes the range-doppler space confronting an SBR intended to be deployed in that environment. There is approximately a 40-fold spread in the value of this parameter, from the Earth to Jupiter's moon Europa. It follows that radar designs that work in one situation may not be at all appropriate if migrated to a different planetary body.

It is often said that radar is "all-weather," but this generalization clearly is not universally true, especially for SBRs. From space, the ionosphere and/or atmosphere may corrupt or even prevent radar propagation. The ionosphere may induce Faraday rotation,[§] thus degrading or destroying the polarization properties of the transmitted and received signals.[4] The Faraday rotation β of a linearly-polarized E-vector is proportional to RM λ^2, where the rotation measure RM is a function of ionospheric electron density. The ionosphere also introduces dispersion and, under certain unfavorable circumstances, effectively cuts off propagation. Thus, for example, it was not possible for the 5-MHz MARSIS radar sounder to probe the Martian surface during daylight hours, because the cutoff frequency under those conditions increased to about 10 MHz. Thus, MARSIS worked as a surface sounder during dark hours and as an ionospheric sounder during daylight hours. More on the MARSIS radar may be found in Section 18.6. The 12-cm wavelength of the Magellan Venus radar (reviewed in Section 18.4) was chosen in response to the trade-off between propagation through Venus' very dense atmosphere (for which longer wavelengths would be better) and synthetic aperture radar system considerations (for which shorter wavelengths would be better). Propagation speed is retarded along the path length from an ocean-viewing altimeter to the Earth by a very small fraction of the speed of light, but sufficient nevertheless to impose range measurement errors of many meters. These errors must be estimated and compensated before the required cm-level accuracy can be achieved, as summarized in Section 18.3.

Comments on Hardware. On a popular American children's television program, one is told that "It is not easy being green." Likewise, it is not easy being

---

[‡] N is the standard symbol for Newton, force, with the units m kg s^{-2}.

[§] The plane of polarization of an EM wave is perturbed by interaction with the magnetic field through which it passes, an effect discovered by Faraday.

a radar, especially aboard a spacecraft. Because a radar by definition must transmit, its near-field radiation is a potential threat to all other instruments and subsystems of the host spacecraft and its payload. Once having got past the risk (or paranoia) of near-field radiation, normal spacecraft design principles kick in. The first order considerations on hardware that are characteristic of (and essentially unique to) the space environment include radiation and energetic particles, vibration (especially during the launch phase), harsh and contrasting thermal environments, often challenging mass limitations, and a premium on payload power. In Earth orbit and, perhaps surprisingly, also at the Moon, a space-based radar must comply with internationally agreed spectral allocations. These limit both the available bands and the bandwidths, which impact system design and may constrain certain performance objectives, such as resolution.

Since it is expensive to get to orbit, SBRs, like other space-based systems, must be designed to minimize mass and to maximize efficiency and longevity. Mass, power, and lifetime emerge as driving themes that dictate conservative design, generous margins during system implementation, and redundancy often realized by dual-string hardware in most if not all subsystems other than the antenna.

Organization of the Chapter. The chapter sections are organized by measurement theme, taken in a broad sense. These themes are Earth-orbiting synthetic aperture radar (SAR), radar altimetry (which in the SBR context almost always implies observation of the Earth's oceans and larger surface water bodies), planetary radars (where "planet" includes planetoids such as large moons), scatterometers (whose data associate a geophysical parameter such as oceanic wind speed to the observed calibrated radar backscatter properties of the illuminated surface), and sounders (which includes both atmospheric and subsurface radar systems). Each section includes an overview of all relevant SBRs, noting key turning points or watershed innovations in the theme's history. Selected examples have been chosen for more in-depth review. Online web sites are suggested for each cited instrument; these web sites were accessible at the date of publication of this book.*

18.2 SYNTHETIC APERTURE RADAR (SAR)

In its most general form, an imaging radar is a device designed to provide a two-dimensional portrayal of the radar backscatter returning from the field illuminated in range and azimuth. Space-based microwave imagers are synthetic aperture radars (with the exception of certain early Soviet ocean-observing real-aperture systems). As with all imaging systems, SAR image products are rated according to their resolution, where "higher" is "better." Higher resolution always implies wider bandwidth in both range and azimuth. Azimuth bandwidth derives from the doppler signatures set up by the motion of the radar with respect to the illuminated field. Resolution by itself is not sufficient to determine the image quality of importance to most applications.

* The field of space-based remote sensing radars is subject to rapid change. This chapter provides the view from the early 21st century. Readers are encouraged to seek currently topical information through online resources. Keywords, such as the names of the mission, nationality, and radar, are usually sufficient to locate several references. Alert: Not all online resources are accurate; the reader is advised to seek out more than one and to verify information by cross-comparison.

SAR images are degraded by a multiplicative self-noise known as *speckle*, which is a direct consequence of the coherence required by the radar-processor combination to form the synthetic aperture and the resulting enhanced resolution. Speckle can be reduced only through supplemental incoherent processing, multi-looking in SAR jargon. Additional looks require proportionally more bandwidth. It follows that large two-dimensional bandwidth (range and azimuth) is the driving requirement for this class of radar.

Space-based SAR systems have motivated fruitful specializations in quantitative applications in a wide variety of areas, comprehensively reviewed in the *Principles and Applications of Imaging Radar*.[5] Topics such as SpotSAR, ScanSAR, polarimetry, and interferometry that have influenced radar system and mission design are outlined in closing paragraphs of this section.

Flight Systems. The concept of SAR, introduced in 1951 by Carl Wiley,[6] was reduced to practice in subsequent years, first through simulations, then through airborne proof-of-concept systems.[7] Quill, the first space-based SAR (see Table 18.2), was launched only a decade later. (That was a remarkable achievement, considering that in the modern era, it often takes nearly 20 years to go from concept to launch of a new SAR, even though the principles and technology for these SBRs are by now well established.) Quill was rudimentary, but did succeed in generating data sufficient to form images. The nominal 100-m resolution was spectacular for its time, given that the best resolution that could be expected from an otherwise comparable real-aperture SBR would be on the order of kilometers. Nevertheless, the results did not meet the needs of the sponsor, and so the notional second and third missions were never launched. Quill was the only American SBR whose data were optically recorded onboard, eventually returned to Earth by ejected capsule, and then collected by an airborne retrieval maneuver.

TABLE 18.2 Synthetic Aperture Radars (Earth-viewing)

Satellite/SAR	URL	Country	Launch	Res (m)	Band	Polarization
Quill	1	USA	1964	(>100 m)	X	
Seasat	2	USA	1978	25	L	HH
SIR A; B	3	USA	1981; ‘84	40; ~25	L	HH
SIR C	4	USA; G, I	1994, ‘94	~30	L&C; X	Various to quad; HH
Kosmos 1870	5	USSR	1987	15–30	S	HH
Almaz	6	USSR	1991	15–30	S	HH
ERS-1	7	ESA	1991	25	C	VV
J-ERS-1	8	Japan	1992	30	L	HH
RADARSAT-1	9	Canada	1995	8, 25, 50, 100	C	HH
ERS-2	10	ESA	1995	25	C	VV
Priroda	11	Russia/Ukraine	1996	50	S, L	HH, VV
SRTM	12	USA; G, I	2000	~30	C, X	HH, VV
ENVISAT	13	ESA	2002	10, 30, 150,1000	C	VV or HH, dual
IGS-1B	14	Japan	2003+	1, +	X	Multimode
PALSAR	15	Japan	2006	2.5–100	L	Various to Quad
JianBing-5	16	China	2006	3–20	L	Multi-polarimetric
TerraSAR-X	17	Germany	2007	1, 3, 15	X	Various
RADARSAT-2	18	Canada	2007	1, 3, 25, 100	C	Various to Quad
COSMO	19	Italy	2007	1, 3, 25, 100	X	Multi-polarimetric
TecSAR	20	Israel	2007	1–8	X	Multimode

TABLE 18.2 Synthetic Aperture Radars (Earth-viewing) (*Continued*)

Satellite/SAR	URL	Country	Launch	Res (m)	Band	Polarization
Kondor-E	21	Russia	2007	1, +	S	Multimode
HJ-1-C	22	China	2007	1, +	S	Multimode
SAR-Lupe	23	Germany	2007	0.12, +	X	Multimode
Arkon-2	24	Russia	2008	1–50	S, L, P	Multimode
RISAT	25	India	2008	1–50	C	Various to Quad
Tandem-X	26	Germany	2009	1, 3, 15	X	Various to Quad
Radarsat-C	27	Canada	–	1, +	C	Various to Quad
MAPSAR	28	Brazil/Germany	–	3–20	L	Single, dual, quad
Sentinel-1	29	Europe	–	4–80	C (L)	Various to Quad

1. http://www.skyrocket.de/space/index_frame.htm?http://www.skyrocket.de/space/sat_mil_usa.htm
2. http://www.astronautix.com/craft/seasat.htm
3. http://directory.eoportal.org/pres_SIRShuttleImagingRadarMissions.html
4. http://southport.jpl.nasa.gov/
5. http://www.astronautix.com/craft/almazt.htm
6. http://www.russianspaceweb.com/almazt.html
7. http://earth.esa.int/ers/
8. http://www.nasda.go.jp/projects/sat/jers1/index_e.html
9. http://www.space.gc.ca/asc/eng/satellites/radarsat1
10. http://en.wikipedia.org/wiki/Space-Based_Radar
11. http://www.astronautix.com/craft/priroda.htm
12. http://www2.jpl.nasa.gov/srtm/
13. http://envisat.esa.int/object/index.cfm?fobjectid=3772
14. http://www.space.com/spacenews/archive03/spyarch_040903.html
15. http://www.eorc.jaxa.jp/ALOS/about/palsar.htm
16. http://www.sinodefence.com/strategic/spacecraft/jianbing5.asp
17. http://www.caf.dlr.de/tsx/start_en.htm
18. http://www.space.gc.ca/asc/eng/satellites/radarsat2/inf_over.asp
19. http://directory.eoportal.org/pres_COSMOSkyMedConstellationof4SARSatellites.html
20. http://www.iai.co.il/Default.aspx?docID=32812&FolderID=14469&lang=en&res=0&pos=0
21. http://www.npomash.ru/space/en/space1.htm
22. http://www.eohandbook.com/eohb05/pdfs/missions_alphabetical.pdf#search=%22China%20HJ-1C%20satellite%20radar%22
23. http://directory.eoportal.org/pres_SARLupeConstellation.html
24. http://industry.esa.int/ATTACHEMENTS/A112/nfm2005_04.pdf#search=%22Japan%20space%20radar%20IGS-R1%22
25. http://directory.eoportal.org/info_RISATRadarImagingSatellite.html
26. http://directory.eoportal.org/info_TanDEMXTerraSARXaddonforDigitalElevationMeasurement.html
27. http://www.mdacorporation.com/news/pr/pr2006031301.htm
28. http://elib.dlr.de/43957/
29. http://www.gmes.info/

Seasat. The generally acknowledged space-based synthetic aperture radar pioneer was the 1978 Seasat SAR (Figure 18-1). That L-band system[8] persists as the design paradigm for Earth-observing space-based SARs. Seasat illustrates several characteristics of many civilian Earth-observing SARs, including the size and aspect ratio of the antenna (10.74 m by 2.1 m), its relatively steep angle of incidence (~22°), its swath width (100 km), and use of a linear FM (chirp) modulated pulse waveform (634:1 compression ratio). Seasat SAR's average radiated power was relatively small (55 W) although its peak power was appreciable (1 kW). The antenna was passive, consisting of eight flat microstrip panels, radiating and receiving HH polarization. There was no onboard recorder, so data were downlinked, but only, of course, when in radio visibility of one of the four ground stations (in the United States, Canada, and the United Kingdom) that were equipped to receive the data. Telemetry was

analog (20 MHz, offset video), to be converted on the ground to either optical media (transparent film strips) or digitized (5-bit quantization). Imagery was generated (but not immediately†) by either optical or digital processing methods. The Seasat satellite suffered a massive short circuit in its primary power system (in the solar panel slip-ring assembly), which ended its mission in October 1978 after only three months of operation.

SIR SAR Series. The Shuttle imaging radars, SIR-A (which flew on the first Shuttle that carried a science payload), SIR-B, and SIR-C/X-SAR were essentially technology (and science) demonstrations missions,[2,9] each lasting about one week or less. In sequence, these radars had increasingly more capability in terms of incidence, frequency, and polarization. SIR-C/X-SAR operated at three bands: C and L (USA) and X (contributed through an international partnership with Germany and Italy). Data from SIR-C continues in demand, thanks to its fully polarimetric multiband coverage of a wide variety of scenes.[10] The Shuttle Radar Terrain Mapper[11] (SRTM) was an outgrowth of SIR-C, using a small receiving antenna (at X and C band) mounted on a 60-m long extendable strut, to collect simultaneous backscatter subsequently processed into topographic maps. This was the first demonstration of single-pass space-based interferometric SAR capability (InSAR). Data were collected over a very large fraction of the global landmass.

Kosmos. Although not classified in the strict sense, Kosmos came to view largely after the launch of its technical successor, Almaz (Russian and Arabic for *diamond in the rough*). Preceding these Soviet SARs were a series of real-aperture radars, known as *Okean* (and by other names). Almaz was a very interesting radar, in that it provided unique S-band SBR imagery of the Earth's surface. The technology behind this and related Russian systems continued to be developed, culminating in the Kondor-E radar and, through a bilateral technology exchange program, the HJ-1-C SAR of China.[12] Both of these radars use the 6-m parabolic reflector developed originally for Priroda, the final module for the Russian MIR complex. Travers, the radar aboard Priroda, was an experimental demonstration mission. There are at least three members of the X-band national asset SAR fleet, including TecSAR[13] (Israel) and the IGS-R series (Japan). The first IGS-R was launched in 2003. The 2005 launch of a sequel failed.

ERS-1 and ERS-2. The C-band SARs from the European Space Agency (ESA) introduced operational space-based SAR capability.[14] Since the launch of the first European Radar Satellite (ERS-1), ESA SARs have maintained a continuous record of outstanding performance. ERS-1 and its followers have had onboard data recording capability. ERS-1 and ERS-2 had in common ~25-m resolution at 4 looks, 100-km swath width, and ~23° incidence, all in the pattern of the Seasat SAR. The two ERS SARs were virtually identical. ERS-1 and ERS-2 enjoyed about one year of simultaneous operation, organized by ESA to chase each other so that their joint repeat visit was only one day. This lead to the accumulation of a unique data set, especially valuable for coherent change detection measurements[15] (two-pass InSAR). The ERS SAR instrument was a subset of a combined radar group, the Active Microwave Instrument (AMI), comprised of a scatterometer and a wind/wave mini-SAR mode, as well as the SAR itself.

† The first digitally processed images from the Seasat SAR required 40 hours for a quarter-frame (50 km square) on a mainframe computer.

(The scatterometer mode is described in Section 18.5.) The wind/wave mode, based on the SAR, was designed to grab 10-km square snapshot images of the ocean surface, which could be processed (onboard or later in a ground facility) to estimate wind speed and direction from the radar brightness and wave pattern in the data. This mode was motivated by a requirement to gather SAR-quality data over a large area within tight data volume limits. One consequence of the ocean-viewing requirements was that VV‡ emerged as the favored polarization, since vertically polarized oceanic backscatter tends to be larger at C band at steep incidence§ than horizontally polarized oceanic backscatter. ERS-1 and -2 also carried radar altimeters (Section 18.3). The ESA series of satellites use midday sun-synchronous orbits because their payloads include optical sensors. The resulting half-orbit eclipse contributes to their per-orbit SAR operational limit of ten minutes.

J-ERS-1. The SAR aboard the Japanese Earth Resources Satellite[14] (J-ERS-1) was similar to Seasat, being at L band, with an H-polarized 11.9 m by 2.4 m antenna. Its 38° incident angle,* which is larger than that of Seasat, was enabled by its lower altitude and larger antenna aperture. Its 1.1 kW peak power was comparable to that of Seasat. One consequence of the larger antenna vertical dimension is the 75-km swath, which is narrower than that of Seasat. J-ERS-1 was crippled by antenna and connector problems, so that its sensitivity (noise-equivalent sigma-zero NEqσ^0) was only –14 dB, 6 dB shy of its –20 dB design. As a result, J-ERS-1 image products were far noisier than those of Seasat, whose sensitivity was –23 dB. In spite of that handicap, however, J-ERS-1 data (at 23-cm wavelength) provided the first synoptic coverage of the tropical forests of Brazil and neighboring countries. These data established an early "gold standard" for tropical forest surveys, since L band is so much better suited to this application than shorter wavelengths. J-ERS-1 had 18-m resolution at 3 looks, about 50% "better" than Seasat, based on comparison of their respective SAR image quality factors (Section 18.4). J-ERS-1 operated for eight years, considerably longer than its two-year design life.

RADARSAT-1 Shown in Figure 18.2, RADARSAT-1 marked a major milestone in space-based SAR. It was the first system to offer the user a choice of resolutions, incident angles, and swath widths.[17] The evolution of these characteristics merits a brief review. The Canadian requirements spanned a variety of applications, from oceanic surveillance (vessels and oil platforms as well as sea state), land- and sea-ice, agriculture, and forests, among many others. The sea-ice application had high priority, which drove the choice of polarization. Horizontal was chosen, because that

‡ It is common in radar remote sensing that their polarizations are abbreviated as an alphabetic pair, in this case indicating vertical polarization on both transmit and receive.

§ Space-based SARs tend to look toward the surface at angles that are much closer to vertical than do airborne SARs. The most commonly used terminology is incidence (the angle at the illuminated surface between the local vertical and the incoming illumination). The angle of incidence is the complement of the grazing angle, the customary nomenclature for airborne radars. Incidence differs from elevation (the angle at the spacecraft between vertical and the direction from the spacecraft to the scene), where the difference is due to the Earth's curvature.

* Alert! The Japanese SARs usually specify the look angle as the elevation angle at the spacecraft, although calling that angle "incidence." Thus, most of the literature cites 35° for J-ERS-1's incidence, which can lead to confusion if the value of this parameter is important in a given application. The same caveat applies to PALSAR, described in a subsequent paragraph.

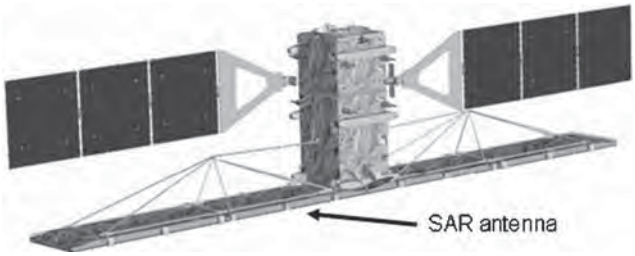


FIGURE 18.2 RADARSAT-2 is outwardly similar to its predecessor. The solar panels are parallel to the along-track axis of the antenna, indicative of a dawn-dusk sun-synchronous orbit. (*Courtesy of the Canadian Space Agency*)

offered an advantage when attempting to distinguish new ice from calm sea at C band and moderate incidence. The applications diversity implied a span of preferred incidence from ~20° to more than 50°. In response, the antenna was required to steer its beam in elevation, which evolved into seven different standard elevation beam patterns, electronically steered and shaped. As might be expected, no applications theme would accept degraded resolution. To satisfy that demand, the radar design adopted three bandwidths, so that the nominal ground-range resolution would be maintained at ~25 m over all baseline incident angles. As a corollary, the wider bandwidth at shallower incidence would result in finer range resolution. The antenna design was based on horizontal runs of 32 waveguides, each center-fed. Ferrite phase-shifters, one for each waveguide, controlled the transmit and receive beam shape and boresight in elevation. The electronic beam-selection enabled RADARSAT-1 to incorporate ScanSAR.† Thus, the resolution options for RADARSAT-1 range from 8 m by 8 m (single look) 45-km swaths (fine mode), to 100 m by 100 m (8 looks), 510 km swaths (ScanSAR Wide). Incidence ranges from 10° to 60°, including the extended modes. There are seven standard modes, each having its own elevation beam; these modes have nominally 25 m by 28 m (4 looks), 100-km swaths. The NEqσ^0 is –20 dB or better and mode-dependent.

By the end of 2006, RADARSAT-1 had completed its 11th year, logging more than 60,000 orbits and collecting enough data to map the entire surface of the Earth an equivalent of 130 times. The Canadian Ice Service relies on RADARSAT-1 data for its routine operations, which require more than 3000 frames of data per year. Since the SAR is the only RADARSAT-1 payload instrument, a sun-synchronous, dawn-dusk orbit was chosen to maximize illumination of the solar panels, which allows 20 minutes of SAR operation per orbit. RADARSAT-1 looks to the right side of the orbit plane, which gives it access to the Canadian Arctic up to the North Pole. Twice during its mission, the spacecraft was yawed 180° for several weeks, which enabled full coverage of Antarctica. The resulting data were merged to yield the first high-resolution imagery of the whole continent,[18] and over several regions, repeat-pass coverage supported interferometric SAR measurement of Antarctic glacial flow rates. RADARSAT-2 is an enhanced version[19] of RADARSAT-1. Table 18.3 charts a summary of its many modes.[20]

† Discussed later in this section.

TABLE 18.3 Rsat-2 Modes

Mode	Width W	Access	Resolution	Looks	$1/Q_{SAR}$	$W\ Q_{SAR}$
Selective polarization	*km*	*km*	*m, Rg × Az*	*Rg × Az*		
Standard	100	250–750	25 × 26	1 × 4	162	0.62
Wide	150	250–650	30 × 26	1 × 4	195	0.77
Fine	50	525–750	8 × 8	1 × 1	64	0.78
ScanSAR Wide	500	250 –750	100 × 100	4 × 2	1250	0.40
ScanSAR Narrow	300	250–720	50 × 50	2 × 2	625	0.48
Single polarization						
Low incidence	170	125–300	40 × 26	1 × 4	260	0.65
High incidence	70	750–1000	18 × 26	1 × 4	117	0.60
Polarimetry						
Standard QuadPol	25	250–600	25 × 8	1 × 4	50	0.50
Fine QuadPol	25	400–600	9 × 8	1 × 1	72	0.35
Selective Single Pol						
Multi-look Fine	50	400–750	8 × 8	2 × 2	64	3.12
Ultra-Fine	20	400–550	3 × 3	1 × 1	9	2.2
Experimental						
MODEX (GMTI)						
Very High Resolution			3 × 1	1 × 1		

ENVISAT. The advanced SAR[21] (ASAR) aboard ESA's Envisat takes RADARSAT-1's versatility one step further, with the addition of two polarizations, H and V, on either transmit or receive. Its 10 m by 1.3 m active array antenna comprises 320 transmit/receive modules. In other regards, its modes reflect the basic RADARSAT-1 design (due, in large part, to the participation by key radar specialists from Canada—a member of ESA's Earth Observation Program Board—on the ASAR conceptual design team). The transmit and receive polarizations are independent of each other, so that at full resolution, the polarization choices are HH, VV, or HV. Note that this dual-polarization mode is actually *alternating*-polarization, in which the polarization states are switched between transmissions or receptions, but the pulse-repetition frequency is not increased. Such a "dual-polarized" pair, by definition, cannot be mutually coherent since it corresponds to interleaved samples of the backscatter, time-multiplexed at less than the Nyquist rate. As a consequence, the phase difference between complex samples at each pixel is not available. The ASAR alternating polarization mode provides dual-polarized images that are similar to those that have been available from many airborne systems for decades, although not available from an SBR (except for SIR-B and SIR-C) until ENVISAT.

PALSAR. In contrast to the incoherent polarimetric options of ASAR, the Phased-Array L-band Synthetic Aperture Radar (PALSAR) aboard Japan's ALOS[‡] (launched in January 2006) includes full quadrature[§] polarimetry.[22,23] The PALSAR modes include standard single-polarization mapping modes, a ScanSAR mode, a variety of dual- and quadrature-polarized modes, and experimental modes, including SpotSAR. Early mission calibration and validation studies verified that the radar is performing as intended. More on PALSAR appears later in this section.

[‡] Advanced Land Observation Satellite, JAXA, Japan.

[§] Discussed later in this section.

JianBing-5. Known alternatively as Remote Sensing Satellite-1, Jianbing-5 is China's first synthetic aperture radar mission. The spacecraft mass is 2700 kg, launched into a sun-synchronous orbit at ~600 km altitude. Mission objectives include polarimetric diversity and interferometry. The two baseline multi-look resolutions are 3 m (40-km swath) and 20 m (100-km swath), over a variety of incident angles. Its antenna is an active phased array.

TerraSAR-X. TerraSAR-X is the first civilian dedicated space-based SAR at X band. Its 4.8 m by 0.8 m antenna is a two-dimensional active array of 384 T/R modules. It has a variety of modes, from ScanSAR (15-m resolution over 100-km swaths) to SpotSAR (1-m resolution over a 5-km by 10-km image frame). Its strip mapping mode is baselined at 3-m resolution across 30-km swaths. The array is partitioned in the along-track direction, which can be exercised in a two-aperture along-track interferometric mode for GMTI experiments, among other applications. Full quadrature polarization is one of the mode options. The only payload instrument is the radar, so the spacecraft is designed for a sun-synchronous dawn-dusk orbit, 11-day repeat. Several years after the launch of TerraSAR-X, it will be joined by a companion, Tandem-X, which is meant to be a functional copy.[24] These two will co-orbit as a closely coordinated pair to support a variety of bistatic and interferometric applications.

Other Space-based SARs. Space-based SAR continues to undergo considerable expansion on an international front. This section provides an overview of programs known (at the time of publication of this Handbook) to have progressed through Phase A, which is evidence of funding that is sufficiently serious that most are likely to culminate in the launch and operation of a flight system. The architecture of these emerging systems is dominated by one of two antenna paradigms: active two-dimensional phased arrays or reflectors. Most missions are multimode in resolution (hence swath and coverage), and polarimetry (ranging from incoherent dual-polarization to full-quadrature polarimetry). At least four of these initiatives imply several satellites, either in constellation or in series. All together, something in excess of 20 new SARs are being launched in the first decade of the 21st century, from at least eight different countries. Highlights of several of these systems are summarized in the following paragraphs.

COSMO-SkyMed. Italy has a series of four COSMO-SkyMed X-band SAR satellites. The COSMO SARs have multipolarization active phased-array antennas that support a variety of modes including 1-m SpotSAR, strip map, ScanSAR, and 500-km wide swath. The COSMO bus is based on the design built and flight-proven for Canada's RADARSAT-2.

TecSAR. The first space-based SAR from Israel, TecSAR is a featured element of their national satellite technology development program.[25] Launched by India, TecSAR's nominal strip-map mode is multi-look 3-m resolution at X band. Major additional objectives include large area coverage as well as high resolution, which imply ScanSAR and SpotSAR modes, respectively. The 1-m SpotSAR resolution constrains the design, one consequence of which is the symmetrical body-mounted 3-m diameter umbrella-style mesh reflector (0.5 kg) driven by one of ten feed horns slightly offset from the focal point. This unity-aspect ratio antenna is a notable departure from the rectangular antennas so typical of most space-based SAR precedents.

FIGURE 18.3 TecSAR features a symmetrical reflector antenna, in contrast to the classic high aspect-ratio "billboard" style pioneered by Seasat. (*Courtesy of IAI, Israel*)

ScanSAR coverage (8-m resolution, 40-km swath, to 20-m resolution, 100-km swath, incidence-dependent) is executed by sequential selection of the appropriate feeds. The feed horns are H- or V-polarized, so that polarimetric diversity is also supported by this arrangement. Mission design includes spacecraft steering to support squinted aspects relative to the reference side-looking orthogonal viewing. The high-power stage is comprised of ten channeled TWTAs,* eight of which are required in combination, leaving two as redundant backup. The radar mass is 100 kg (including the reflector and feeds); satellite dry mass including the radar is ~300 kg, which at the time set the standard as the smallest SAR satellite in Earth orbit. The power subsystem delivers up to 1.6 kW during imaging operations. The spacecraft orbit is 143° inclination, ~550 km altitude, 36-day repeat period. Since this is not a sun-synchronous orbit and since the antenna pattern (to first order) is symmetric, the spacecraft may be rotated about the radar's line-of-sight to help sustain near-full illumination of the two solar panel arrays in all sun-orbit plane configurations. Radar data are recorded in a 256 Gbit solid-state unit following 6-to-3 block-floating-point quantization. Spacecraft design life is five years.

HJ-1-C. HJ-1-C is the first of five SAR satellites in China's small satellite program[24] announced in 2002 for environment and disaster monitoring. The architecture of these systems is closely parallel to Russia's *Kondor-E* series, since they are both outgrowths of a bilateral development program based on the architecture of the Priroda S band radar. The HJ-1-C's antenna is a reflector that has an effective aperture of 6 m by 2.8 m following deployment. In strip-map mode, the HJ-1-C has a multi-look 5-m resolution over a 40-km swath, and 20-m resolution in ScanSAR mode at 100-km swath width. SpotSAR mode is supported by controlled yaw maneuvers of the spacecraft. All of the spacecraft in the series (five radars and six optical systems) use sun-synchronous orbits at ~500 km altitude. The radar's mass is ~200 kg.

SAR-Lupe. Germany has five identical X-band satellites, distributed in three 500-km high orbits, inclined at about 80°. The aspect ratio of their 3.3 m by 2.7 m

* Traveling-wave tube amplifier. See Chapter 10 for more discussion.

antennas suggests that the dominant objective is fine resolution, which necessitates a relatively narrow range swath. Published specifications note that the intended resolution in the slip-SAR† mode is 0.12 m over a 5 km by 5 km image frame. These innovative SBRs are relatively small, at least by Earth-observing satellite SAR standards. Their 770 kg mass (total dry mass of the spacecraft and the radar) is less than the mass of RADARSAT-2's antenna! The SAR-Lupe design also is cost-conscious, based on rigid (nondeployed) reflector antennas ("borrowed" from a commercial communications satellite production line) that are inherently more efficient and less massive than an active array. The radar electronics are directly descended from a commercial product line.

RISAT. The Radar Imaging Satellite, or RISAT, is India's first space-based SAR,[27] following extensive programs in optical remote sensing satellites and airborne imaging radar development. RISAT's deployable antenna (6 m by 2 m) is an active phased array comprised of 288 C-band (5.35 Ghz) TR modules, each capable of 10 W peak power. The average output power (200 W) requires an average input dc power of 3.1 kW. Each TR module is connected to separate distribution networks feeding H- and V-polarized elements, which support polarization diversity as well as elevation beam steering. There are two parallel receive channels (dedicated to the H- and V-polarized antenna elements). RISAT has five modes, each of which may operate at a variety of incident angles. The modes are: fine-resolution strip-map-1 (3-m resolution, 30-km swath, dual-polarization); fine-resolution strip-map-2 (12 m at 30 km, quadrature-polarization); medium resolution ScanSAR (25 m at 120 km, dual-polarization); coarse resolution ScanSAR (50 m at 240 km, dual-polarization); and high-resolution SpotSAR (better than 2 m, 10-km square image frame, dual-polarization). The SpotSAR mode requires yaw/pitch steering of the spacecraft to ±13°. Coverage to either side of the ground track requires a roll maneuver of the spacecraft to direct the antenna pattern to the opposite side of nadir, an approach that is similar to that of RADARSAT-2. The variety of resolutions is supported by four bandwidths (225 MHz, 75 MHz, 37.5 MHz, and 18.75 MHz) through a programmable digital chirp generator. The received data are downconverted to baseband, digitized to 8 bits (I and Q), and quantized (by block-floating point means) to fewer bits (6 to 2), at the user's option within mode-dependent limits. All subsystems (save for the antenna) are dual-redundant. Nominal PRF is 3250 Hz ± 450 Hz. Data rates span 142 Mbits/s to 1478 Mbits/s, depending on mode. The on-orbit mass of the spacecraft will be ~1750 kg, of which the SAR payload (including the antenna) claims ~950 kg. Onboard data storage capacity is 240 Gbits; downlink maximum data rate is 640 Mbits/s (X band, dual circularly polarized). The RISAT orbit is sun-synchronous, dawn-dusk, at ~609 km altitude, and 13-day revisit period.

MAPSAR. The Multi-Application Purpose SAR[28] is a joint Brazilian-German enterprise aimed primarily at assessing and monitoring Brazil's natural resources. After several years of trade studies, L band was selected (1.3 Ghz). The radar is built around a near-symmetrical reflector (7.5 m by 5 m) with ten feeds, offset from the focal point so that the beam can be electronically scanned in elevation. Spatial resolution and swath span 3 m to 20 m, and 20 km to 55 km, respectively, mode-dependent.

† Slip-SAR is a modified ScanSAR mode, in which the antenna pattern is dragged along the surface at a slower rate than in the conventional strip-map mode. The result is larger doppler bandwidth, hence enhanced azimuth resolution, and also a wider imaged area than a pure SpotSAR can support.

The high-resolution specification results from an external constraint, the maximum 85 MHz bandwidth stipulated by international spectral allocation agreements. There is provision for all polarimetric configurations. The sun-synchronous orbit will be controlled to assure reliable 37-day revisit baselines to support interferometry. Nominal instrument mass is 280 kg.

PALSAR. Although based on J-ERS-1 heritage, PALSAR[29] is an excellent early example of multimode space-based SARs. PALSAR's 132 modes (Figure 18.4) include standard single-polarization mapping, ScanSAR, a variety of dual- and quadrature-polarized modes,‡ and experimental or demonstration modes including SpotSAR. The L band (23 cm) baseline frequency (1270 MHz) has two bandwidths, 28 MHz (fine-beam single-polarization mode), and 14 MHz (dual-, quad-pol, and ScanSAR modes). The span of mid-swath incident angles is (7.9° –60°). (Note that the Japanese usually cite these angles as "off-nadir" (9.9°–50.8°), which refers to their elevation with respect to vertical in spacecraft coordinates, rather than their intersection relative to a local vertical at the Earth's mean oblate spheroidal surface.)

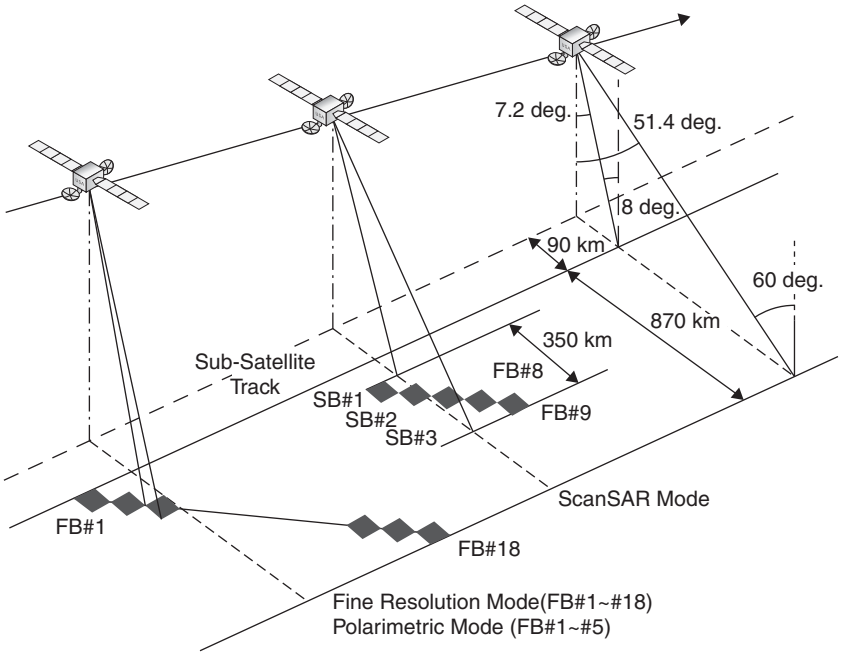


FIGURE 18.4 Overview of PALSAR's viewing geometries. Each of these beam positions supports a variety of polarization combinations, leading to a very large number of available modes. The solar panel is orthogonal to the orbit plane, indicative of a midday sun-synchronous orbit. (*Diagram courtesy of JAXA, Japan*)

‡ The several forms of polarimetric diversity common in remote sensing systems are reviewed later in this section.

Modern space-based SARs owe their large mode variety to active electronic-steered antenna arrays. PALSAR's antenna consists of 80 solid-state T/R modules distributed on four panels whose deployed dimensions are 3.1 m by 8.9 m, vertical and horizontal, respectively. Spacecraft velocity and antenna length dictate a lower bound on pulse repetition frequency, which for PALSAR spans 1.5 kHz–2.5 kHz, depending on mode. Peak transmitted power is ~2 kW, twice that of the Seasat SAR. The resulting sensitivity (noise-equivalent sigma-zero) is very good, –30 dB or better for most modes. PALSAR's host ALOS spacecraft is yaw-steered to maintain azimuth antenna boresight pointing at zero doppler, which increases the likelihood of pass-to-pass interferometric coherence and simplifies (somewhat) SAR image formation processing. The nominal data rate in most modes is 240 Mbit/s, which is downlinked via Japan's Data Relay Test Satellite (DRTS). The ScanSAR mode requires only half that rate, 120 Mbit/s, which can be downlinked directly to ground stations. ALOS has an onboard 96 Gbyte solid-state recorder to buffer data output from the radar, as well as from the rest of the payload.

PALSAR's voluminous variety of modes is a curse as well as a blessing. Mission management must cope with data collection in each of these modes, as well as phasing the resulting data burden to share the communications link with the two other high data-rate instruments in the ALOS payload. The standard policy for the early years of PALSAR's on-orbit scheduling is to focus on six "default" modes—four "operational" and two "semi-operational." The resulting terms are (1) fixed "standard" off-nadir angle of 41.5° for a great majority of the data takes; (2) polarization options to be single-pol HH and dual-pol HH+HV; (3) quad-pol at 21.5° (off-nadir) for R&D demonstrations over preselected "supersites"; and (4) five-beam ScanSAR in HH polarization. In addition, the following constraints apply: only one mode is exercised during one 46-day repeat cycle; preferred operations are during the hours of darkness in the ascending passes for most modes, excepting lower data-rate ScanSAR data takes during descending passes (in coordination with the optical sensors) and also excepting extraordinary InSAR and marine applications at nonstandard incidence; and recurrent repeat-pass coverage of selected sites in groups of eight or more 46-day cycles in support of InSAR objectives.

Space-based SAR Design Issues. The options for space-based SAR design[30,31] are more limited than for airborne systems, due primarily to the constraints imposed by viable orbits, including especially sensor velocity, radar range (R), and system cost. The following paragraphs review the major themes.

PRF Constraints. The rules that govern the pulse repetition frequency (PRF) of a space-based SAR are the same as those applicable to airborne systems, although they play out rather differently. The fundamental requirement is that the PRF f_p be sufficiently high to sample unambiguously the doppler spectrum of width B_{Dop}, and also sufficiently low so that there is time between transmissions to receive the data backscattered from the intended swath of slant range (time-domain) width T_R. Thus

$$B_{\text{Dop}} < f_P < 1 / T_R \tag{18.2}$$

In practice, sufficient margin must be included in both the upper and lower limits to account for the length of the transmitted pulse, and for the fact that neither the doppler spectrum nor the antenna's elevation pattern have sharp cutoffs.

The lower bound often is recast as

$$B_{\text{Dop}} = \frac{2\beta V_{\text{SC}}}{\lambda} = \frac{2V_{\text{SC}}}{D_{\text{Az}}} < f_P \tag{18.3}$$

which states that the PRF must be sufficiently high such that there are two transmissions per antenna aperture length D_{Az} as the radar moves along its trajectory. (Here V_{SC} is spacecraft velocity along its orbit, β is the width[§] of the azimuth antenna pattern, and λ is the radar wavelength.) This form is intuitively satisfying, since the single-look azimuth resolution is approximately one-half of the antenna's along-track width, hence this inequality requires that each resolved along-track range line must be sampled at least once. Usually the PRF lower bound is set so that there is a margin of 25% or more with respect to this constraint.

In an airborne SAR, the PRF constraint is derived to satisfy the doppler bandwidth, from which follows the maximum range that the radar can operate without introducing ambiguities. However, by default, the minimum range for a space-based SAR is its orbital altitude, usually 600 km or higher. The typical slant range to the intended scene may be 800 km and more. Thus, the upper bound on the PRF should not be set by the range to the scene, but rather by the range width of the area to be imaged. As a consequence, the resulting high PRF will generate a sequence of pulses at any moment that are distributed between the radar and the scene. The space between pulses must be larger than the intended swath width. For example, in certain modes RADARSAT-1 generates seven pulses "in flight" simultaneously. At the beginning of such a data collection, backscatter from the intended scene would arrive only after the seventh pulse had been transmitted.

In many airborne systems, the PRF is chosen to be rather higher than the limit imposed by the range to the intended target space. In such cases, the extra PRF contributes to improved SNR, but at the cost of increased average data rate. Data rate can be decreased by "pre-summing"—coherently adding n adjacent returns. The consequence, of course, is that the effective PRF is reduced by the same factor n. This practice is seldom acceptable for space-borne SARs, because it leads to azimuth ambiguities unless the doppler spectrum is limited prior to PRF reduction.

Ambiguities. The PRF generates a two-dimensionally sampled space when the data are decomposed into "slow time" (in the azimuth direction) and "fast time" (in the range direction). In azimuth, the PRF creates aliased versions of the data illuminated by the main beam of the antenna. The spectra of these aliases are located at multiples of the PRF to either side of the doppler centroid of the main beam. Of course, when sampled they are folded back into the Nyquist passband. These aliases are azimuth ambiguities, which are suppressed—hence not visible—in a well-designed system supported by a well-tuned processor. When they do appear, azimuth ambiguities are relatively easy to identify (Figure 18.5) because they are weaker (ghost) duplicates of image features that were collected through the main beam and, therefore, at earlier or later positions along the image strip. The azimuth shift of the ambiguities relative to the central image is an integral multiple of $\Delta X = R\lambda f_P / (2V_{\text{SC}})$, which is the spatial

[§] For a uniformly illuminated aperture, the −3 dB beamwidth is $\beta = 0.88\ \lambda / D_{\text{Az}}$. It is customary for analysis of SAR systems to approximate this expression by $\beta \sim \lambda / D_{\text{Az}}$ and to interpret β as the width of the rectangle that has the same peak value and area as the antenna pattern.

offset corresponding to the PRF. Azimuth ambiguities, especially from point targets, have the same doppler FM rate as scatterers within the main beam, and so their focus is preserved through the processor.

In the range direction, one of the consequences of having many pulses in flight at once is that there are echoes from several different ranges that arrive back at the radar at the same relative delay within the range gate as the reflections from the intended swath. If these extra echoes are sufficiently strong, the resulting image artifacts are range ambiguities. Range ambiguities are not as easy to identify as azimuth ambiguities, because they arise from ranges outside of the nominal swath, hence not otherwise imaged. Range ambiguities by definition arise from ranges that are different from those for which the processor is set, so that range-ambiguous point targets tend to be defocused.

The principal means of suppressing ambiguities is to confine the main beam of the antenna so that the potential sources of azimuth or range ambiguities are not illuminated, or at least are illuminated only very weakly. This requirement imposes a minimum area constraint on the SAR's antenna. The lower and upper bound PRF constraints of B_{Dop} and $1/T_R$ lead to

$$D_{\text{El}} D_{\text{Az}} > 4RV_{\text{SC}}(\lambda / c)\tan\theta_{\text{Inc}} \tag{18.4}$$

where the antenna area is the product of its length D_{Az} and height D_{El}, and θ_{Inc} is the mean incident angle in the imaged swath. The range-velocity product in this expression is determined by the parameters peculiar to the particular planet (or moon) about which the SAR is to operate (Table 18.1). As a consequence, an antenna of area 1 m^2 at the Moon would have be 10 m^2 for the same radar at Mars and nearly 40 m^2 for operation in Earth orbit.

It always is tempting to push this constraint, since most space-based SARs use as much of the ambiguity-free space as can be tolerated. In practice, the antenna area usually is chosen to be at least a factor of two larger than this minimum would suggest. Ambiguities are proportional to the strength of the offending backscatter, and as such, they contribute to the multiplicative noise* ratio (MNR) of the system. Antenna sidelobes and ambiguities are further suppressed by appropriate weighting in the processor. The trade-off is lower MNR, at the expense of broader impulse response width (IRW). Typical designs provide ~20% excess range and doppler bandwidth relative to those implied by the required range and azimuth resolutions to accommodate such weighting.

Nadir Return. A potentially troublesome range ambiguity arises from the surface directly below the spacecraft. This nadir return is always relatively strong, particularly if there are specular backscattering components. Since any realistic antenna pattern has a non-zero sidelobe directed toward nadir, the resulting reflection could appear in the image. The main strategy is to avoid the ambiguity by choosing the PRF so that the nadir return arrives at the same time that the radar is transmitting. This timing places a further constraint on the pulse repetition frequency. It turns out that the nadir return may not be avoidable if other constraints override the available PRF options. This occurs, for example, if the driving requirement is ScanSAR, which has its own set of constraints on PRF.

* Multiplicative noise, a standard specification in synthetic aperture radar, includes unwanted contributions such as ambiguities and quantization noise that are proportional to the strength of the received signal.

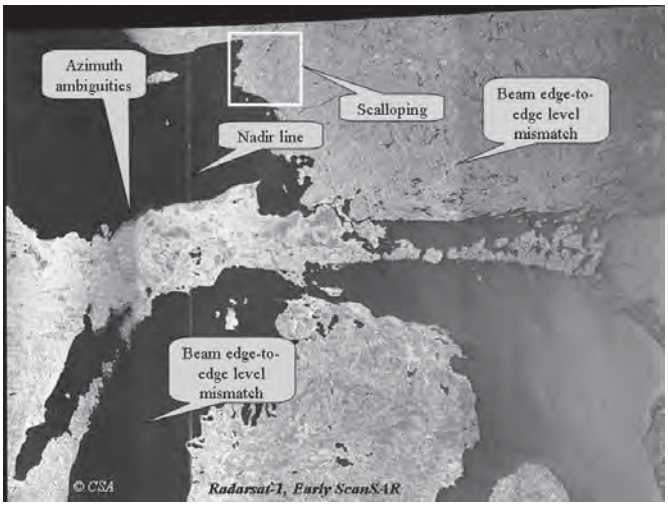


FIGURE 18.5 Example of artifacts as they appeared in early RADARSAT-1 four sub-swath ScanSAR imagery. The direction of flight is vertical in this presentation, near-edge range at the left of the frame. (*Courtesy of the Canadian Space Agency and the Canada Centre for Remote Sensing*)

Antennas and Transmitters. Antennas for the pioneering space-based SAR systems were exclusively passive, such as the L-band patch array on Seasat or the slotted waveguide arrays used on ERS-1/2 and Almaz. The first exception to this rule was RADARSAT-1, designed around an array of 32 horizontal slotted waveguides, each of which was center-fed through a phase shifter so that the elevation beam pattern could be selected and shaped electronically. More ambitious (read "massive and expensive") systems tend to use two-dimensionally active electronic scanned arrays (ESAs). These are populated by transmit/receive (T/R) modules, often incorporating two polarizations (H and V). Examples include RADARSAT-2, ENVISAT's ASAR, PALSAR, COSMO, TerraSAR-X, and RISAT. It is argued that active arrays set the stage for "graceful failure" since the loss of a few T/R elements would have little effect on the net performance of the system. In parallel, an alternative SAR antenna paradigm stresses simplicity and lower mass (and lower cost) over two-dimensional electronic beam steering. These are reflectors, of which China's HJ-1-C, Germany's SAR-Lupe, Israel's TecSAR, and Brazil's MapSAR are good examples. If a reflector is driven by multiple feeds, then one may still effect beam steering, although with rather less beam shape variety and control than through an ESA.

Space-based radar transmitters naturally fall into two classes, intimately coupled with the antenna's architecture. If the antenna is active, then the transmitter (as well as the front end of the receiver) is distributed over the array. In this case, several hundred T/R elements, of a few watts peak power each, add up to many hundreds of peak radiated power. Phase control of the elements is a critical parameter, usually requiring adaptive temperature compensation to assure coherency of the radiated wavefront. The alternative is almost always limited to traveling wave tube amplifiers (TWTAs), although recent developments in high-power solid-state devices is influencing SBR SAR design. Radars built around TWTAs have established impressive longevity records, witness RADARSAT-1 and ERS-2, both of which remained in operation for more than ten years.

Data Rate. Data rate is proportional to the product of pulse repetition frequency f_P, the number of range samples N_R (which is proportional to slant-range swath, plus the uncompressed pulse length, and inversely proportional to range resolution), the number of quantization bits N_S retained in each sample of the data, and a factor of 2 that accounts for the in-phase (I) and quadrature (Q) components since both the amplitude and the phase of the signal stream are required. Once resolution and swath are established, the number of bits per sample is the only parameter open to choice. Excellent results may be obtained for N_S, as small as 2 bits (per I and per Q) by adapting the quantization thresholds to the mean signal level. (Refer to Section 18.4 for more on this topic.) Quantization noise, which is larger for fewer bits, is proportional to signal strength; hence it is a factor in the MNR budget. Exacting applications such as interferometry are better served by more bits per sample, subject to the rate and volume limits of the data handling subsystems.

Processing. Although similar in principle to airborne systems, processing for space-based SARs differs in several key regards. The highlights are reviewed here. For more complete coverage, see the standard references.[30,32] A simplified introduction to the key parameters may be helpful.[33]

The natural starting point is the range equation, from which several properties emerge that are unique to the spherical geometry of radars in orbit. In the narrow-beam side-looking case, and neglecting the effects of Earth rotation, the radar-to-target range variation generates a phase time-history over a synthetic aperture length of T seconds from a point reflector at minimum slant range R_0

$$\Theta(t) = -\frac{4\pi}{\lambda}\left(R_0 + V_{\mathrm{SC}}V_{\mathrm{Beam}}\frac{t^2}{2R_0}\right) \tag{18.5}$$

where V_{Beam} is the velocity on the surface of the illuminating footprint of the azimuth antenna pattern. The time derivative of the phase yields the scatterer's doppler history

$$f_D(t) = -\frac{2}{\lambda}\frac{V_{\mathrm{SC}}V_{\mathrm{Beam}}}{R_0}t \tag{18.6}$$

in which the FM rate is proportional to the effective velocity $V_{\mathrm{Eff}} = (V_{\mathrm{SC}}V_{\mathrm{Beam}})^{1/2}$. This is in contrast to the airborne case for which the FM rate of the azimuth doppler modulation is proportional to the square of the aircraft velocity. Why the difference? Rather than a straight line, which is the baseline situation for an airborne system, the synthetic aperture of a space-borne SAR is formed along an arc. This imposes a small but significant increase in the effective length of the synthetic aperture and also modifies the FM rate. As a consequence, the nominal single-look azimuth resolution from a space-based SAR is $r_{\mathrm{Az}} = (V_{\mathrm{Beam}}/V_{\mathrm{SC}})D_{\mathrm{Az}}/2$ rather than the famous "one half of the aperture length" of airborne SARs. Note that V_{Beam} is always smaller than the spacecraft velocity, and decreases with increasing spacecraft altitude and incident angle. Alert: In certain SAR processing literature, the effective velocity for the orbital case is denoted "radar velocity," a rather misleading and inappropriate term.

The average data rate from space-based SARs in Earth orbit is on the order of 100 Mb/s, (megabits per second) with higher resolution and polarization diversity systems generating several times that. The drivers are the spacecraft velocity (~7.5 km/s), range resolution, and swath width. Many users would like to have immediate access to processed data, which leads to the question of onboard processing. Space-based

SAR systems as a rule do not implement image formation onboard for several reasons, including high data rates. Processing from raw SAR data to imagery increases data volume substantially, thus increasing the downlink data transfer burden. Perhaps more persuasive, once processed into detected images, the options are limited for specialized application-specific post processing.

In contrast to airborne systems, the dynamics of most spacecraft are such that no motion compensation is required for space-based SARs (unless extremely fine resolution is to be generated). However, determining the doppler centroid of the azimuth spectrum with respect to zero doppler emerges as a dominant issue. If the azimuth antenna boresight is perfectly orthogonal to the inertial orbit plane, the Earth's rotation imposes a doppler shift onto the received data.[30] From low Earth orbit (LEO), the magnitude of this shift is on the order of 3°, and (to first order) varies sinusoidally with latitude, with maximum magnitude at the equator, and with zero at the extreme N and S latitudes. The effect can be offset by yaw-steering the spacecraft so that the antenna's boresight is always directed toward zero doppler. This orients the vertical plane of the azimuth boresight to be orthogonal to the nadir track on the surface, rather than to the orbit plane. Whereas SAR data from both arrangements can be processed into satisfactory imagery, more demanding applications such as radar interferometry are better served by yaw-steered systems. Yaw steering imposes negligible additional demands on the spacecraft attitude control system, since it requires maneuvers of only ±3° over each orbital period. Note that a SAR looks down as well as to the side, so that vertical velocity components in the satellite's orbit also lead to doppler shifts in the data. In principle, a doppler shift from a vertical velocity component also could be offset by attitude adjustments of the spacecraft, although this strategy is not in general practice. Doppler centroid estimation is a central function implemented in all processing algorithms for SBR SAR data.

Data Products. The notional data product from a space-based SAR is imagery, usually visualized as a black-and-white mapping of the scene illuminated by the radar. By definition, the digital numbers at each pixel in such an image array are real and non-negative. In theory, these numbers correspond to the magnitude-squared of the focused and detected backscattered field. In practice, most image products (such as those from the European Space Agency's SARs) use magnitude, because the resulting imagery has an acceptable appearance and the size of the data file is smaller than if in $|\text{magnitude}|^2$. If several data sets over the same scene are combined, often these are individually color coded, leading to multicolor image products. Each constituent data set might be from a different polarization, wavelength, or observation time.

It is common parlance to refer to imagery as mappings of the normalized backscatter power $\sigma^0(x,y)$. Alert: In fact, this is seldom *if ever* true. Backscattered power is proportional to $|\text{magnitude}|^2$, not magnitude. Hence, the user must assure that the data are indeed magnitude-squared before applying tools, such as speckle filters, that are designed for σ^0 dimensionality. Second, σ^0 implies that the data are calibrated, not only with respect to the radiometric parameters of the radar and processor, but also with respect to the local incident angle at the pixel location (x,y). Although data from ERS-1/2 are corrected to account for the mean incidence within the imaged swath, there is no attempt to correct for slopes locally within the swath to the pixel level. An alternative is to denote the (magnitude-squared) digital numbers as β^0, which indicates simply radar power per pixel.[31] This has become standard practice with RADARSAT-1 data, for example.

Standard image products usually are "multi-look." In jargon common in the space-based SAR field, "looks" refer to statistically independent versions of the same scene. When these are added together, the net result is to reduce the speckle noise, while reinforcing the imaged features. Each such look is formed from a spectral band that does not overlap spectra corresponding to the other looks. Thus, for a given bandwidth, increasing the number of looks reduces speckle, but at the cost of compromising resolution. (More on this trade-off may be found in Section 18.4.)

In contrast to conventional post-detection imagery, focused SAR data may be presented as single-look complex (SLC) products. These data retain the full resolution of the radar, and most important, retain the relative phase of the backscattered field. By definition, SLC files are in amplitude and phase, often represented as an array of in-phase (I) and quadrature (Q) signed number pairs, at each pixel. SLC data are required for SAR interferometry, polarimetry, and coherent change detection.

Pushing Ambiguity Limits. The baseline single-look SAR azimuth resolution is proportional to one over the doppler bandwidth generated by the azimuth beamwidth of the side-looking antenna. The corresponding length of the synthetic aperture is equivalent to the along-track spread of the antenna pattern, which, of course, is proportional to range. Let this be the canonical case. Azimuth resolution may be sharpened only by increasing doppler bandwidth, which can be done in one of two ways: increasing the antenna's beamwidth or increasing the spread of aspect angles within which the antenna illuminates a given portion of the scene. The latter is the basis for Spotlight SAR,† in which the antenna is steered to dwell on the intended area as the radar passes, thus creating a wider total doppler bandwidth (and a longer synthetic aperture). The trade-off is that adjacent areas along-track may not be imaged at all. (Increasing azimuth resolution by broadening the antenna pattern, either by reducing the aperture length or spoiling the beam, has the disadvantage of reducing the antenna's gain, which usually is not desirable for space-based SARs. In addition, the PRF must be larger than the instantaneous doppler bandwidth, thus reducing the allowable unambiguous range swath.)

Going the other direction—smaller doppler bandwidth—leads to more coarse azimuth resolution. The doppler bandwidth of the original signal history from a given backscatterer may be reduced by the simple expedient of generating a shorter synthetic aperture than the canonical case. This logic leads to the "burst mode," which figures prominently in two forms in space-based SARs. Burst mode along a single imaged swath implies a reduced data rate, which may be necessary to meet the stringent data-rate requirements confronting planetary or lunar missions. Alternatively, the intervals between bursts may be used to illuminate several different range swaths, thus expanding the area that may be imaged unambiguously. This is the principle behind the ScanSAR mode.

Ambiguity Space Trade-offs. It is easy to show that these resolution and coverage options are consistent with the principles that govern range and azimuth ambiguities. The fundamental rule is that the image space (illuminated by the antenna) must be "underspread" if ambiguities are to be avoided.[34] The underspread condition is that

$$T_R B_{\text{Dop}} < 1 \tag{18.7}$$

† Alternatively known as SpotSAR.

where T_R is the (slant) range swath depth of the antenna pattern and B_{Dop} is the corresponding doppler bandwidth. To first order, the azimuth resolution is given by

$$r_{\text{Az}} = \frac{\beta R N_L}{T_{\text{Az}} B_{\text{Dop}}} \tag{18.8}$$

where β is the azimuth beamwidth, R is slant range, N_L is the number of looks (here assumed to be taken in the azimuth domain), and (T_{Az} B_{Dop}) is the azimuth time-bandwidth product, comprised of the illumination time of the target, and its doppler bandwidth. Substituting for doppler bandwidth in the underspread condition leads to the constraint

$$\frac{T_R \beta R N_L}{T_{\text{Az}} r_{\text{Az}}} < 1 \tag{18.9}$$

which shows how resolution and target illumination time may be traded against each other while still respecting the fundamental ambiguity constraint. The following four cases are important in practice.

Strip Map. The standard approach is strip mapping, in which the ambiguity-free space is (nearly) filled with the allowable swath, resolution, and number of looks. Of course, (azimuth) resolution can be increased within the allowed space at no loss of swath width if the number of looks is decreased in proportion.

SpotSAR.[35,36] If (azimuth) resolution is the driving objective, then r_{Az} may be reduced as long as the integration time T_{Az} is increased in proportion. Given that the beamwidth of the antenna is fixed, the integration time can be increased only if the beam is pointed to maintain illumination of the desired target, much as a spotlight from a moving vehicle dwells on an area of interest. The usual consequence of fine resolution in SpotSAR mode is smaller range swath and an azimuth size that is limited above by the width of the antenna's footprint. The required steering rate is relatively slow, typically a few degrees over a few seconds, implemented either by moving the antenna beam or by yaw-steering the spacecraft. Note that the radar's PRF needs to be only above the Nyquist limit set by the antenna beamwidth, not by the total span of doppler frequencies collected. Range resolution is improved in the usual way, by increasing the radar's bandwidth, for which the Stretch technique is often helpful. A variation of this mode (SlipSAR) is to drag the antenna footprint rather than to stare at one area, thus trading poorer azimuth resolution than a pure SpotSAR would deliver to gain increased azimuthal coverage.

How far can this be pushed? Under the assumption that a target's scattering is coherent over a full 180° sector, it can be shown[31] that the ultimate azimuth resolution is $\lambda/4$. By the way, such phenomenal results have been approached in the field of seismology.

Burst Mode.[37] If average data rate is the driving consideration, then the integration time T_{Az} may be reduced below the canonical limit set by the azimuth antenna beamwidth. This is accomplished by turning off the transmitter after enough pulses have been collected to satisfy the azimuth resolution requirement. Each such burst has an instantaneous doppler corresponding to the antenna beamwidth (which determines the Nyquist PRF limit), but a shorter synthetic aperture length. Burst mode is standard

operating procedure for planetary or lunar radars, for which high image resolution is not required and the spacecraft-to-Earth data link is severely limited. While in burst mode, it is customary to use each burst as a single-look data take, setting the burst repetition frequency so that the desired number of looks is collected during the synthetic aperture length of the azimuth beam pattern. The challenge is to calibrate the antenna pattern such that the framelets from all bursts may be combined to assemble a continuous image along track. Mismatches appear as "scalloping"—systematic brightness modulations at boundaries between each of the framelets.

ScanSAR.[38–40] If swath width is the driving requirement, then azimuth resolution can be traded for range coverage. The trick is to multiplex several burst-mode data sets, where each set of bursts corresponds to a different range sub-swath. In this form of burst mode, the transmitter is always "on"; the burst range sub-swath allocation burden falls onto the antenna. ScanSAR requires rapid elevation beam steering, such as through a phased array (e.g., TerraSAR-X), or selection of one among several range-offset feeds facing a reflector (e.g., HJ-1-C). In addition to suppressing scalloping, good ScanSAR imagery requires that the several range sub-swaths be knitted together such that the crossovers between antenna patterns are not evident. RADARSAT-1 was the first mission to implement (then to perfect) operational ScanSAR, which has been adopted as a standard mode on many space-based SARs. Swath widths achieved are as much as five times the nominal strip-map width, which is considerably larger than the conventional ambiguity constraint would allow. Normally, the trade-off in azimuth resolution is balanced with a corresponding compromise in range resolution, leaving excess range bandwidth that may be converted into looks. Starting from a 100-km swath 25-m × 25-m 4-look system, a reasonable ScanSAR mode could be 500-km swath, 100-m × 100-m × 8-looks, with no increase in average data rate or transmitted power.

Note that all of these ambiguity-space trades start with the canonical case. The trade-offs are relative to that starting position. A radar having an inherently small range swath, for example, would require ScanSAR to expand the range swath to a few tens of km, which may still be much larger than the same radar could cover without ambiguities in a strip-map mode.

Multiple Channels: Interferometry and Polarization. Phase comparison between two or more mutually coherent data sets leads to rich new possibilities, especially including interferometry[15] and polarimetry.[41] This is particularly relevant to space-based SARs, which have been and will continue to be rich sources of quantitative microwave measurements of a wide variety of surface features, enabled by these multichannel capabilities. The following paragraphs provide only a glimpse into these topics; the discussion is meant to whet the appetite of the reader and to provide leads to the voluminous literature.

Interferometry. Interferometry by radar (Figure 18.6) implies measurements that are based on phase differences sensed through two different observations of the same phenomena.[42–45] Phase differences arise from microwave-scale changes that are due either to differential viewing aspect or to elements of motion in the scene. In general terms, the sensitivity of phase measurements depends to first order on (1) the radar's wavelength, (2) the spatial or temporal baseline of the contributing data sets, and (3) the scale of spatial or temporal differential signals. The interferometric baseline increases with range and with platform velocity. It follows that for space-based radars,

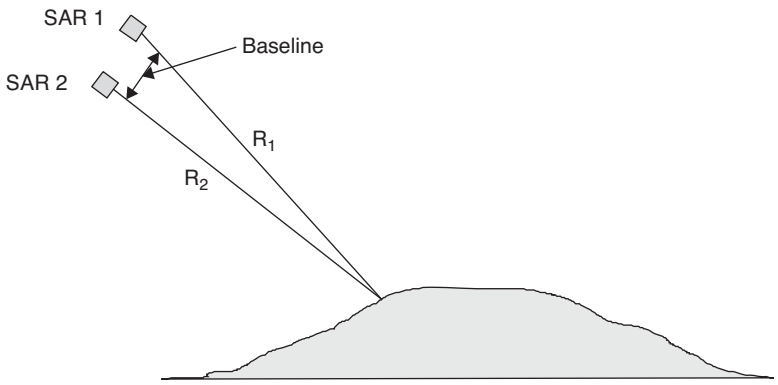


FIGURE 18.6 Radar interferometry expresses the interference created between two mutually coherent backscattered fields. The phase difference interferogram (modulo 2π) corresponds to the relative elevation of the illuminated terrain (*after removal of the systematic slant-range and Earth-curvature signatures*).

conventional interferometric SAR measurements from one spacecraft platform usually are not practical, since the implied spatial or temporal separations of the measurements are much larger than one satellite could support. (An alert reader may note the obvious exception to this rule, SRTM described previously, which mounted its second interferometric antenna on a 60-m extendable strut.)

Satellite orbits tend to be well-known, and follow very similar trajectories over subsequent repeat periods. It follows that the space-based environment offers an attractive alternative: repeat-pass interferometry, originally suggested by Goldstein.[46] If the pass-to-pass observations are separated in the vertical plane, then interferometry leads to relative terrain height estimation.[47] If two observations have a time delay corresponding to the repeat period of the orbit (typically 10 to 45 days), then sub-wavelength movements (in the line-of-sight direction of the radar) may be measured.[48] The technique can be extended to multiple passes with proportionate increase in the temporal baseline, leading to quite remarkable results. Repeat-pass techniques are well suited to mapping topographic relief and to long-term coherent change detection for mapping glacier movement or terrestrial subsidence. If shorter time scales are of interest, to detect moving vehicles, for example (GMTI), then a shorter interferometric baseline is required, which implies two (or more) SBRs in relatively tight co-orbiting formation.

The basic SAR interferometric environment is one in which there is a pair of mutually coherent images that have embedded phases that depend on the details of viewing geometry and scene structure. The coherently combined images are known as an interferogram, which typically contains fringes that express the interactions between the phase structure of the two data sets. Signal processing is designed to estimate these phase differences, and to deduce geophysical parameters of interest from the resulting measurement.[49]

The interferometric signal model is simple in concept. For any neighborhood in the scene, the input signal pair may be described by

$$s_1(t) = \Gamma_1 a \exp[-j\varphi] \tag{18.10}$$

and

$$s_2(t) = \Gamma_2 a \exp[-j\varphi + j\Delta\varphi(r, t_2 - t_1)] \tag{18.11}$$

where the subscripts on the reflectivity Γ suggest that the two signals may be obtained at two different points in time as well as from two different spatial perspectives. The objective is to estimate the relative phase difference $\Delta\varphi$, usually found through the cross-correlation

$$E[s_1(t_1)\overline{s}_2(t_2)] = R_{12}(t_1, t_2) a \exp[-j\Delta\varphi(r, t_2 - t_1)] \tag{18.12}$$

which is presented as it would be calculated using complex image data. Here, $\boldsymbol{E}[.]$ is the expectation (averaging) operator. The phase difference $\Delta\varphi$ may be due either to geometric or to temporal differences between the two observations. Successful interferometry depends on the cross correlation $R_{12}(t_1, t_2)$ of the scattering functions Γ_{t_1} and Γ_{t_2}. The normalized cross-correlation function is the *mutual coherence function*

$$\gamma_{12}(t_1, t_2) = \frac{R_{12}(t_1, t_2)}{\sqrt{E[|s_1|^2]}\sqrt{E[|s_2|^2]}} \tag{18.13}$$

in parallel to that encountered in physical optics.[50] Gamma is a quantitative mapping of the coherence between the two observations. In general, scene coherence decreases with shorter wavelength and longer time between observations.

Mutual coherence is an essential ingredient for radar interferometry. Coherence implies two constraints: spatial and temporal. The spatial constraint applies to the spacing between the orbital passes. Ideally, the radar wavelength projected onto each area of the surface must be the same from both orbits. Since the two orbits are separated, each area is observed at a slightly different incident angle. This implies that the effective wavelength projected onto the ground plane is slightly different for the two cases. Interference is supported only if the range bandwidth of the radar signal is sufficient to span the projected wavelengths, which becomes more demanding as the orbit separation increases.[51] Fortunately, the range pulse has sufficient bandwidth (usually more than 15 MHz) so that mutually coherent range bandwidths can be chosen from the data at the time of processing. Loss of mutual coherence through increasing orbit spacing is known as *baseline decorrelation*.[52] One may show for reasonably level terrain that the upper bound constraint on the difference $\Delta\theta_{\text{Rad}}$ in elevation angle between the two orbits is $\Delta\theta_{\text{Rad}} = \lambda \tan\theta_{\text{Rad}} / 2r_R$, where r_R is slant range resolution, which is inversely proportional to range bandwidth. (*Alert*: In the SAR interferometric literature, it is customary to use elevation angle, defined as the angle between the radar line of sight and the Earth radius vector, as seen from the radar.) For typical large time-bandwidth product signals, the elevation angle constraint implies that correlation of the pair of return signals is maintained for orbital separations on the order of a kilometer. For absolute elevation maps to be derived, however, precise knowledge to the level of one meter is required of the spatial separation between observation tracks.

Temporal coherence applies primarily to the scene. In order for the two signals to act as an interferometric pair, their respective phase structures must be relatively stable over the time interval between satellite observations. In short, there must exist mutual coherence between the two scattering signals, even though they are observed at different times. This requirement is readily satisfied for short interopportunity intervals, such

as from the three-day repeat orbit first used with Seasat data to prove the concept and for stable terrain features, such as unvegetated rocky mountain slopes. It is not necessarily satisfied for scenes, such as ice or vegetation, that may undergo changes in the details of reflection and scattering between observations.

In certain natural and most urban scenes, there will be many corner-reflector-shaped features whose phase remains stable over very long time scales. These so-called permanent (or persistent) scatterers[53] support differential interferometric measurements that may span many repeat visits of the radar, leading to remarkable sensitivity to slow-motion phenomena. For example, multiframe DInSAR‡ analysis of RADARSAT-1 data has led to maps of New Orleans subsidence rates,[54] which vary from ~3 mm/y to more than 15 mm/y, with a sensitivity on the order of 2.5 mm/y.

Any approach to phase difference measurement is subject to the fundamental 2π ambiguity characteristic of phase estimation algorithms.[55,56] In many radar situations, knowledge of the physical constraints of the situation, coupled with phase unwrapping algorithms, is sufficient for the purpose.

Polarimetry. For any given polarization of the transmitted wave, the reflectivity process in general will give rise to a diversity of polarizations in the backscattered wave. To observe these, the radar must be dual-polarized. Similarly, the reflectivity is a function of the polarization of the transmitted wave. Thus, if the scattering function itself is to be fully characterized, two orthogonal polarizations must be transmitted as well. Increasingly, polarimetric diversity is being implemented in space-based SARs.[41,57] Implementation of a fully polarimetric radar always implies greater data channel capacity, more transmitted power, and less range swath coverage. Of course, the antenna—the radar's "polarization gateway"—must be able to receive, and possibly also to transmit, on more than one polarization.

In general terms, there are four options available for polarimetric diversity in a space-based SAR. These are:

(1) **Single (monostatic) polarization** Typical of all dedicated spacecraft SARs (until the launch of ENVISAT) with either HH or VV polarizations. This customary notation indicates horizontal or vertical (linear) polarizations on both transmission and reception for these single-polarized radars.

(2) **Dual polarization** The traditional definition is transmission on one polarization (usually linear, such as H), and reception on the like-polarized and the cross-polarized components (such as H and V). In a traditional dual-polarized radar, the relative phase between the two polarized data sets is discarded. In Earth-oriented remote sensing radars, typical combinations include HH and HV, for example, or HH and VV (which requires two separate transmit polarizations). The ASAR aboard ENVISAT is the first space-based example of this type of polarization diversity. If the four (linear) polarization possibilities are exploited noncoherently, then the backscattering function of the scene may be characterized by the three backscatter coefficients (σ^0_{HH}, σ^0_{VH}, σ^0_{VV}), which, of course, are devoid of phase. (Note that reciprocity implies that $\sigma^0_{HV} = \sigma^0_{VH}$.)

(3) **Coherent dual polarization** A dual-polarized radar that retains the relative phase between the two received polarizations is a significant departure from traditional dual-polarized systems. The modifier "coherent" helps to distinguish

‡ Differential interferometric synthetic aperture radar.

such radars from their more common counterparts described in the preceding paragraph. Coherent dual polarization has not been exploited in orbital SARs (although it is standard practice in Earth-based radar astronomy through facilities such as the Arecibo radar telescope[58]). Experience has shown that there is relatively little added value in the phase between the like-polarized and cross-polarized returns under the condition of H or V transmit polarization. However, an innovative alternative is to transmit circular polarization and receive coherently two orthogonal linearly polarized components. (See Section 18.4 for further discussion.)

(4) **Full or quadrature polarization** This is the richest option because it allows full characterization of the complex matrix of the backscatter at all resolved points in the scene. It has been developed extensively in theory and in practice with data from airborne systems and SIR-C. Japan's PALSAR is the first operational space-based system to incorporate a quad-pol mode.

Primary interest in fully polarimetric radars derives from the enriched scattering observation possibilities revealed through replacement of the scalar form of reflectivity by its complex vector counterpart.[41,57,59] Thus, when either H or V polarizations are incident on a scattering element, both polarizations are backscattered according to

$$\begin{bmatrix} E_H^B \\ E_V^B \end{bmatrix} = \begin{bmatrix} S_{\mathrm{HH}} & S_{\mathrm{HV}} \\ S_{\mathrm{VH}} & S_{\mathrm{VV}} \end{bmatrix} \begin{bmatrix} E_H^T \\ E_V^T \end{bmatrix} \tag{18.14}$$

where the superscript B denotes the field components reflected back toward the radar. The new terms of interest represent the scene's 2×2 scattering matrix, an array of four complex numbers. Each element in this backscattering matrix expresses the magnitude and phase imposed onto the backscattered field (superscript B) in response to the illumination from the transmitted field (superscript T), according to their respective polarizations. Thus, the scattering matrix is a quantitative description of the transformation of polarization state upon reflection, as well as the magnitude and phase of each reflection coefficient. It is often true that the field polarizations are not changed during propagation, the irrotational assumption. In this case, the polarizations of the backscattered waves are equivalent to those that arrive at the radar. This property characterizes most of the polarimetry literature, at least in remote sensing applications, and is reflected in these paragraphs. (The principal exception to this rule is Faraday rotation, which may be a significant factor for longer wavelength systems, such as P band and to a lesser extent L band.)

For irrotational propagation, the portion of the backscattered field captured by the radar is determined by the polarization vector of the receiving antenna, $[E^R]$. The signal voltage v_{rec} entering the system may be written in vector-matrix form as

$$v_{\mathrm{rec}} = \begin{bmatrix} E_H^R & E_V^R \end{bmatrix} \begin{bmatrix} E_H^B \\ E_V^B \end{bmatrix} \tag{18.15}$$

This is the starting point for SAR quadrature polarimetry. Essentially, a quad-pol SAR is managed so that the transmitter generates both orthogonal polarizations. The resulting quad-pol data set can be transformed to represent all possible combinations of transmit and/or receive polarizations. Pitfalls await the unwary, however, including tricky coordinate conventions,[59] an initially confusing variety of "standard" forms for

representing the data, and several alternative methods for data analysis. Once mastered, however, quad-pol SAR data is the undoubted gold standard for quantitative scene characterization by an imaging radar.

A space-based quad-pol capability implies a significant cost. The driving requirement is that the data must be mutually coherent. That is relatively easy for the receiver, which only needs to have two channels that capture simultaneously the phase and amplitude of two orthogonal polarizations of the backscattered field. On the other hand, only one polarization can be transmitted at a time. Illuminating the scene with two polarizations requires the transmitter to be toggled between orthogonal polarization states. This multiplexed transmission scheme implies that the radar's PRF must be doubled to satisfy the minimum Nyquist sampling rate simultaneously for each pair of transmissions. Doubled PRF implies that the average radiated power must be doubled and the unambiguous range swath is halved, both in comparison to the standard case of transmitting only one polarization. Note that the average data rate is the same as the dual-polarized case, since twice as much data are collected by a quad-pol mode for each resolved point in the scene, but the swath width is smaller by a factor of two.

Considerable progress continues to be made in the development of tools for quantitative analysis[60] of polarimetric SAR data. When in combination with interferometric data,[61] the field is known as PolInSAR, for which dedicated specialist meetings are convened frequently. An important methodology is target decomposition,[62] through which specific backscatter classes (such as double-bounce, Bragg, or volume) of a scene may be separated from other types, thence subjected to interferometric analysis. Using such techniques, it is possible to estimate the topography of the surface beneath a vegetated canopy, for example.

Applications. SARs are the largest class of space-based remote-sensing radars, primarily as a result of their practical utility. Many applications are energized by radar's natural ability to operate at night or through cloud, fog, smoke, and haze, and its inherent sensitivity to changes within the scene at wavelength scales. Radar imagery has proven to be valuable for a wide variety of applications, from oceanographic observations (the theme that motivated Seasat) to measurement of millimeter-scale displacement (such as the subsidence of urban areas or the swelling of volcanoes prior to their eruption). Canada's requirement to maintain near-continuous monitoring of its northern and coastal ice is met primarily by thousands of frames of RADARSAT data per year. India is the second largest consumer of space-based imaging radar data, used for agriculture and forest management, and for measuring changes in its alpine glaciers. Countries such as Brazil graced with tropical forests rely on space-based radar imagery to maintain surveillance and to compile annual statistics of deforestation. Since radar imagery is a reliable method of mapping slicks on the ocean's surface, it is the principal means of monitoring oil spills that may result from a grounded tanker or a vessel illegally pumping its bilges in a coastal area. The imaging radar reference[5] cited earlier provides an excellent review of many of these applications.

18.3 ALTIMETERS

In its most general form, an altimeter is a radar device designed to measure the vertical distance between the radar and the surface below. In airborne applications, the resulting "altitude" is a measure of the clearance beneath the aircraft. Whereas the

main objective of a space-based altimeter is also to measure the distance between the radar and the surface, the most common application is determination of the local sea level relative to the Earth's geoid,[§] rather than the height of the spacecraft. The reference for this measurement—the orbital height of the spacecraft—must be known by other means to within a few centimeters. Sea-surface height is a function of many geophysical parameters, such as current flow, an El Niño event, and variations in the ocean's depth. Relatively small changes (on the order of cm) in mean sea-surface height may correspond to substantial differences in the corresponding geophysical parameters. It follows that range measurement *accuracy* and *precision* are the driving requirements for this class of radar. The accuracy of an altimeter's height measurement depends to first order on knowledge of the spacecraft's height along its orbit and on correction of the propagation delays suffered by the radar's round-trip waveform. The precision of an ocean-viewing altimeter is proportional to the radar's range resolution and inversely proportional to the square root of the number of statistically independent measurements (looks) combined for each data point. Ocean-viewing altimeters, in general, have large SNR. Thus, bandwidth and looks become the driving requirements on system design. The emphasis in this section is on altimeter precision.

Sea-surface height measurements have become essential for a wide variety of applications in oceanography, geodesy, geophysics, and climatology.[63] With the exception of near-polar ice, Earth-orbiting oceanographic altimeters have seen relatively little application over nonaquatic surfaces.

A satellite-based altimeter systematically circles the Earth, generating surface height measurements along its nadir track. These measurements accumulate, providing unique synoptic data that have revolutionized our knowledge and understanding of both global and local phenomena, from El Niño to bathymetry. SBR altimeter data also provide measurements of significant wave height and wind speed. Although one might consider altimeters to be relatively simple one-dimensional (range measurement) instruments, their phenomenal accuracy and precision requires elegant microwave implementation and innovative signal processing.

Overview. The surface-height measurement objectives of space-based altimeters can be grouped into four broad categories: large-scale dynamic sea-surface topography, dynamic mesoscale[*] oceanic features, static mesoscale sea-surface topography, and ice—sea ice as well as continental ice sheets. Each of these measurement themes implies narrowed constraints on choice of orbit and on the top-level instrument and mission design. Satellite altimeters dedicated to determining the ocean's large-scale dynamic surface topography are characterized by absolute sea-surface height (SSH) 1-sec-averaged measurement accuracy on the order of centimeters along tracks of more than 1000 km, and orbits that retrace their surface tracks every 10 to 20 days. In contrast, mesoscale missions focus on sea-surface height signals of less than ~300 km in length. Rather than absolute SSH accuracy, these shorter-scale applications require precision sufficient to sustain surface slope measurement accuracy on the order of 1 microradian (one mm sea-level change over a 1-km distance). For geodetic signals that are expressed through static sea-surface topographic variations, orbits are required that generate dense track-to-track spacing. Observation of oceanic and

---

[§] The average sea level in the absence of dynamic perturbations of the surface elevation due to tides and currents.

[*] In the field of oceanography, mesoscale features have scales of several hundred kilometers, as opposed to the much larger basin scale (the North Atlantic ocean, for example).

polar ice sheets requires that the altimeter have robust range and spatial resolution, accuracy, and precision regardless of the non-zero average surface slope in both the along-track and cross-track direction of the continental glaciers. Suitable orbits for ice sheet missions must have near-polar inclination and multiyear relative accuracy of a few centimeters.

Whereas the methodology of these instruments is to determine the distance between the radar and the surface, like any radar, an altimeter actually measures round-trip delay, not distance. At the accuracy required of a space-based oceanographic altimeter, the deceptively simple proportionality of range to delay-time must take into account the small but significant retardation of the radar's microwaves as they propagate. The cm-level SSH accuracy required of these instruments is much smaller than the ranging errors introduced by delays through the ionosphere and the atmosphere. The delays imposed by the ionosphere are a function of frequency. In practice, these can be estimated and then corrected if the altimeter measures round-trip range at two different frequencies. The delays imposed by the atmosphere are comprised of two components: the dry atmosphere and water-vapor. The dry atmosphere component is well-known and stable at long spatial scales; in practice, the resulting delay is compensated by recourse to model predictions. Delays due to water-vapor in the atmosphere are variable down to scales of several hundred kilometers (and much smaller when traversing a storm front). Standard practice is to measure the integrated water vapor contribution in the vertical column below the altimeter by a microwave radiometer, for which two or three frequencies are required.

Measurement errors are dominated by the accuracy of orbit height determination and by the intrinsic precision of the instrument. Figure 18.7 shows a summary history of these factors. The data show that 2-cm instrument accuracy is the state-of-the-art for conventional altimeters. The delay-doppler instrument (see below) would further improve instrument precision[64] to 1 cm.

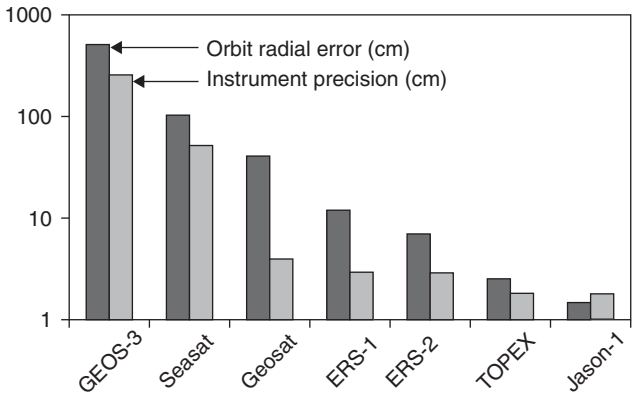


FIGURE 18.7 History of precision orbit determination (POD) and intrinsic instrument precision of the leading radar altimeters of the past 30 years. Vertical axis in cm. Modern POD accuracy relies on GPS and the French DORIS system. Precision is limited by the altimeter's degrees of freedom (incoherent waveform averaging). (*after Dudley Chelton, Oregon State University, personal communication*)

Centimeter-scale range accuracy is supported in the oceanographic application by averaging over the range response of many returns. The range resolution of each return waveform is typically on the order of 0.5 meters. These waveforms are accumulated and averaged pulse-to-pulse, whose shape converges on the flat-surface impulse response[65,66] (Figure 18.8). Sea-surface height (SSH) is derived from the time delay to the midpoint of the waveform's leading edge rise. One thousand or more such waveforms averaged over one second correspond to a mean range estimate whose standard deviation is on the order of centimeters (which, in practice, degrades with increasing significant wave height). One-second averages are standard for operational altimeters, which implies an along-track resolution on the order of 7 kilometers, determined primarily by satellite velocity. Averaging is "the name of the game" in radar altimetry.[67] For example, global data sets from instruments such as TOPEX and Jason-1 have been analyzed to estimate the rate of mean sea-level rise to an accuracy of 1 millimeter per year.

In addition to sea-surface height, the satellite radar altimeter's waveform supports two other oceanographic measurements: significant wave height (SWH) and surface wind speed (WS). Over a quasi-flat sea, a pulse-limited altimeter's idealized mean

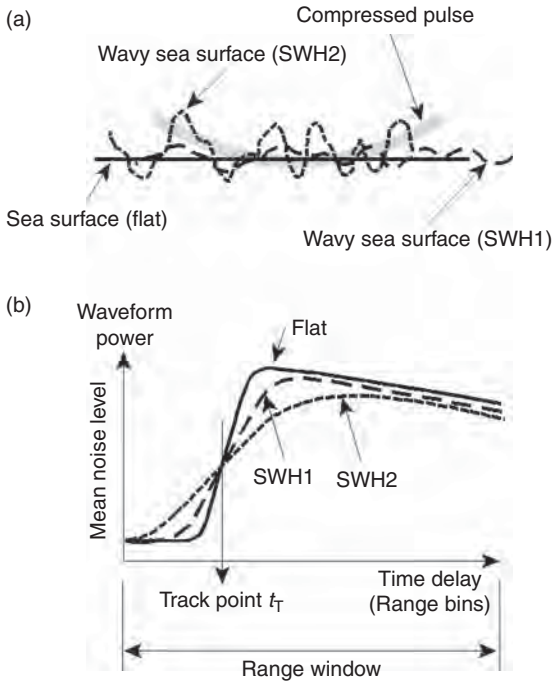


FIGURE 18.8 (*a*) The altimeter's pulse (typically 0.5 m long after compression) sequentially encounters oceanic surface waves (of height up to 20 m or more). (*b*) Sea-surface height (SSH) corresponds to the midpoint of the waveform's leading edge, significant wave height (SWH) to the slope of the leading edge, and wind speed (WS) to the (inverse) backscattered power. The waveforms depicted here are idealized; useful "smoothness" requires 1000 or more incoherently averaged radar returns.

waveform is a step function, whose rise time is equal to the compressed pulse length and whose position on the time-delay axis is determined by the altimeter's height. If the sea surface is modulated by waves, the altimetric depth of the surface increases, which reduces the slope of the waveform's leading edge. Hence, SWH is proportional to the waveform rise time. If the sea surface is disturbed by the wind, the resulting fine-scale roughness decreases the power of the pulse reflected back to the altimeter. Hence, for wind speeds of more than about two knots, WS is inversely related to mean waveform power. In practice, the inflections of the idealized flat-surface response function waveform are softened by the pulse weighting, and the waveform is attenuated over time by the weighting of the antenna pattern.

To extract SWH and WS from waveform data, finely tuned algorithms have been developed and validated against in situ buoy measurements.[63] For example, the TOPEX K_u band altimeter measures SWH to within ±0.5 m up to more than 5.0 m and WS within ±1.5 m/s up to more than 15 m/s. These figures correspond to averages over 1 second, or about 7 km, along the sub-satellite path of the altimeter's footprint, which typically is 3 km–5 km wide, determined by mean sea state.

Flight Systems. Key attributes of satellite radar altimeters are summarized in Table 18.4. Since 1973, oceanic height measurement accuracy has improved, due primarily to more effective means of estimating and correcting systematic errors. Performance also has benefited from innovative onboard hardware and algorithms and more precise determination of the radial component of the orbit. The Jason-1 altimeter represents the state-of-the-art in absolute sea-surface height measurement accuracy (as of the year 2007).

S-193 and GEOS-3. The first satellite radar altimeter was the proof-of-concept S-193 instrument[68] that flew on three Skylab missions. Its objectives were to verify predicted waveform response to wind and waves, to measure the radar cross section of the sea at vertical incidence, to measure inter-pulse correlation properties, and to observe the effect of off-nadir antenna orientation (a scatterometer experiment). Geos-3 provided the first geodetic and geophysical results of significance within the National Geodetic Satellite Program, including the first maps of sea-level variability and the

TABLE 18.4 Altimeters

Spacecraft	Country	Year	Repeat	Inclination	Altitude	Spacing	Band	H2O correction	Accuracy
			days	*degrees*	*km*	*km*			*cm*
Skylab (3)	USA	1973	No	~48	435	n/a	K_u	–	50 m
GEOS-3	USA	1975–8	No	115	845	~60	K_u	–	50
Seasat	USA	1978	~17, 3	108	800	160, 900	K_u	Yes	20
Geosat	USA	1985–1989	*GM*, 17.05	108	800	~5, 160	K_u	–	10
ERS-1	ESA	1991–1996	3, 35, 176	98.5	785	900,80,15	K_u	Yes	7
TOPEX / Poseidon	USA/ France	1992–2005	9.916	66	1336	315	C, K_u K_u	Yes	2 5
ERS-2	ESA	1995–	35	98.5	785	80	K_u	Yes	7
GFO	USA	1998–	17.05	108	800	160	K_u	Yes	5
Jason-1	France	2001–	9.916	66	1336	315	C, K_u	Yes	1.5
Envisat	ESA	2002–	35	98.5	785	80	S, Ku	Yes	7
Jason-2	France	2008	9.916	66	1336	315	C, Ku	Yes	1.5
Altika-3	India (Fr)	2009	35	98.5	785	80	Ka	Yes	1.8
CryoSat-2	ESA	2009	369	92	720	n/a	Ku	–	5
Sentinel-3	Europe	2010	35	98.5	785	80	C, Ku	Yes	5

marine geoid.[69] Geos-3 and the S-193 altimeters used conventional pulse-compression techniques. As suggested in the table, neither of these two early altimeters included a water-vapor radiometer, and each used only one frequency, so that they had no in-built means to correct for ionospheric or atmospheric propagation delays.

Seasat's Altimeter. Seasat's was the first to use full deramp† pulse compression,[70] which opened the way for the very small range resolution required for many oceanographic applications. The deramp technique (described below) has been adopted by all radar altimeters since then. Seasat was designed to measure global ocean dynamic topography, as well as wave height and surface wind speed.

Geosat. This altimeter's design[71] was patterned closely after that of the Seasat altimeter. Geosat was a U.S. Navy military satellite whose primary mission was to map the Earth's marine geoid to then-unprecedented accuracy, for which a non-repeat orbit was required. Since its public release in 1995, the data set from the first 18-month geodetic mission has become the backbone of the global bathymetric chart that is the industry standard.[72,73] Geosat's secondary mission was to observe dynamic mesoscale oceanographic phenomena, for which it was maneuvered into an exact repeat orbit (period 17.05 days).[74] Geosat's geodetic mission and exact-repeat mission are known as GM and ERM, respectively. As a spacecraft, Geosat was one of the few Earth-observing missions to rely on the purely passive gravity-gradient means of attitude control,[74] as evidenced by the extended (vertical) boom in Figure 18.9. The attitude was stable to less than 1°, to which the altimeter's pulse-limited range measurement was robust.

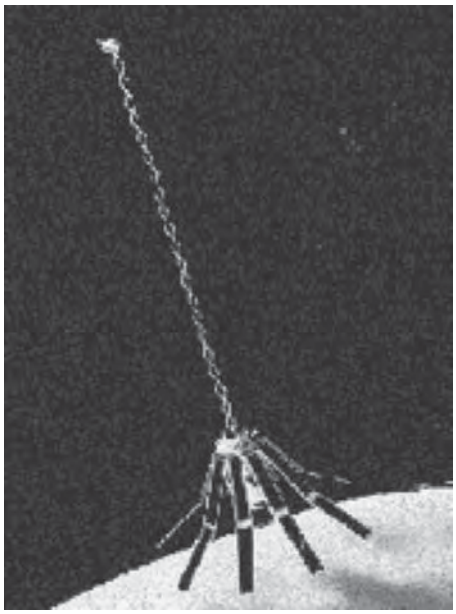


FIGURE 18.9 The Geosat radar altimeter: The nadir-directed antenna (a reflector) is hidden inside of the solar arrays. This spacecraft maintained verticality by gravity-gradient (hence the long boom and counterweight). The data were impervious to randomly oriented yaw angles (about the vertical axis). (*Image courtesy of Johns Hopkins University Applied Physics Laboratory*)

Geosat Follow-On (GFO). GFO was designed to replicate as much as possible the Geosat exact-repeat mission, in support of operational requirements for the U.S. Navy. GFO represents the current state-of-the-art in small dedicated radar altimeter missions. It includes a dual-frequency water-vapor radiometer (WVR) at 22 and 37 GHz, data from which are used to reduce the corresponding propagation uncertainty to 1.9 cm.

† Full deramp (or simply deramp) is standard terminology in space-based radar altimetry. It is known more commonly to most radar engineers as the Stretch technique.

The radar's mass is 45 kg (including full redundancy and the WVR); its prime power consumption is less than 100 W. Total spacecraft dry mass‡ is about 300 kg.

TOPEX/Poseidon. In the late 1980s, program planning for satellite radar altimeter missions split into two themes, determined by the relative priority of their measurements. If the altimeter were the prime payload instrument, then the orbit and mission design could be optimized accordingly. This theme was followed by TOPEX/Poseidon (T/P), a joint United States (NASA) and French (CNES) mission. TOPEX was designed to measure and map the dynamic ocean topography with sufficient accuracy to determine large-scale circulation patterns.[75] TOPEX' most famous contribution is early observation and near-real-time monitoring of El Niño events, whose height signature over the equatorial eastern Pacific ocean typically is an increase on the order of 10–20 cm with respect to the mean. Poseidon, contributed by France, was a small proof-of-concept instrument that had a solid-state transmitter. Poseidon was the precursor of the Jason altimeters and the SIRAL instrument aboard CryoSat.

The T/P orbit repeat period was chosen carefully to satisfy adequate observation of the dominant aliased tidal constituents. All solar tidal constituents would be ambiguous with other height signals if the repeat period were an integral number of days. For T/P, the time of day for each subsequent observation slips by about two hours. The T/P repeat pass footprint location accuracy is better than ±1 km, a requirement that is bounded by the cross-track gradient of the oceanic geoid. The T/P instrument package included a three-frequency radiometer. TOPEX was the first altimeter to use two frequencies (time-multiplexed) to estimate and compensate for propagation delays[76] imposed by ionospheric electrons. Designed for an initial three-year mission, subsequently stretched to five years, T/P provided valuable data for an impressive thirteen years. T/P was formally decommissioned in December 2005.

As is true for most radars, the received waveform produced by an individual pulse from TOPEX was corrupted by coherent self-noise known as speckle. The standard deviation of speckle is reduced by summing (averaging) many statistically independent waveforms together. Statistical independence between sequential returns observed by a radar altimeter depends primarily on the radar pulse repetition rate, the antenna size, the spacecraft velocity, and on the sea surface conditions.[67] The antenna was a 1.5-m reflector that served both altimeter bands and also the radiometer. Selected parameters associated with the altimeter's design are listed in Table 18.5. The pulse-to-pulse

TABLE 18.5 TOPEX Parameters

Parameter	Value	Units
LFM rate	3.125	MHz/μs
Pulse duration	102.4	μs
Pulse BW (radiated)	320	MHz
Time × bandwidth	32768	(dimensionless)
Pulse resolution	0.469	m
Carrier (K_u band)	13.6	GHz
Carrier (C band)	5.2	GHz
IF frequency	500	MHz
Stretch bandwidth	3	MHz
Range time span	400	ns

‡ Total mass, not including consumables such as fuel for propulsion or attitude control.

statistical independence requirement evaluated for the TOPEX indicates that the maximum PRF should be 2.5 kHz, yet it was 4.5 kHz in practice. The pulse rate above the threshold improved the additive SNR, but did not contribute to speckle reduction. The PRF statistical independence limit decreases with increasing significant wave height.

Jason-1. Jason-1 followed in the footsteps of TOPEX, figuratively and literally. After the launch of Jason-1 into the T/P orbit, TOPEX was maneuvered into a "tandem" phasing so that the measurements of the two altimeters could be cross-calibrated. Jason-2 is essentially identical to Jason-1.

ERS-1, ERS-2, and ENVISAT. If the altimeter is not the primary payload, then the resulting mission and orbit are likely to be determined by other requirements, which may compromise altimetry. The European Space Agency's satellite altimeters on ERS-1 and ERS-2, as well as the advanced radar altimeter[21] RA-2 on ESA's ENVISAT, are of second priority with respect to the other instruments on their respective spacecraft. Their sun-synchronous orbits are less than optimum for most altimetric applications, due primarily to the fact that four of the eight dominant tidal constituents are sun-synchronous.[63] These orbits are also at lower altitudes than the T/P orbit, which implies that orbit maintenance maneuvers must be more frequent, thus compromising precision orbit determination. During a portion of its mission, the orbit of ERS-1 was rephased to a long repeat period (176 days). That long repeat period generated a relatively dense surface sampling grid useful for estimating sea ice cover, geodesy, and bathymetry. The ERS-2 mission did not vary its repeat period. The consequence of these orbit properties is that the resulting data are not well suited to measuring the annual rate of sea-level rise, which is a key climate-related variable.

CryoSat. This was the first Earth Explorer Opportunity Missions, which was part of the European Space Agency's Living Planet Programme. The mission concept[77] was selected in 1999 and subsequently launched in October 2005. Unfortunately, the launch vehicle malfunctioned. ESA and its member states authorized a replacement. The CryoSat orbit has a high-inclination (92°) and a long-repeat period (369 days, with a 30-day subcycle), designed to provide dense interlocking coverage over the polar regions. Its aim is to study possible climate variability and trends by determining the variations in thickness of the Earth's continental ice sheets and marine sea ice cover. CryoSat is described in more detail in a subsequent section.

AltiKa. AltiKa differs from other ocean-viewing altimeters in this section, due primarily to its use of K_a band (35.75 GHz) rather than K_u band. The first instrument (from France) is part of the payload on India's Oceansat-3. AltiKa[78] is single-frequency since at K_a band the retardation due to the ionosphere is sufficiently small that it does not have to be measured and compensated. However, the ~0.84 cm wavelength is vulnerable to atmospheric moisture; it is predicted that as much as 10% of the data will be compromised by rain. The 33 kg instrument requires an input power of 80 W. The offset-fed reflector antenna is 1 m in diameter, resulting in a beamwidth less than half that of its K_u band counterparts. Several advantages are claimed for the smaller beamwidth, including operation closer to land. On the other hand, the narrower beam implies that the waveform is more sensitive to spacecraft attitude errors. AltiKa's 500 MHz bandwidth leads to a pulse-limited footprint about 30% smaller than usual.

The PRF is 4 kHz, approximately twice that of most conventional altimeters, slightly larger than the pulse-to-pulse statistical independence condition of 3.75 kHz.

Orbit Considerations. Given an arbitrarily good radar altimeter, its orbit becomes the dominant factor that may limit sea-surface height measurement accuracy.[79] Orbit selection for an ocean altimeter requires consideration of the impact of an orbit's inclination, repeat period, and altitude. For example, if the objective is absolute sea-surface height accuracy over larger spatial scales and long time scales, then a higher altitude orbit having a relatively moderate prograde inclination, and relatively short non-sun-synchronous repeat period is the only sensible starting point.

An altimeter's revisit period is ten days or more, in contrast to tides with approximately 1 or 2 cycles per day driven primarily by lunar and solar gravity. As a result, all tidal signals sensed by an altimeter are undersampled. Altimetric data retain the resulting aliases, which over the course of a year or so can be identified, quantified, and calibrated out. An altimeter's orbit must be chosen so that the tidal aliases do not get confused with signals of geophysical interest.

The T/P Orbit. The state-of-the-art (at least in accuracy and large-scale circulation studies) is Jason-1, operating in the orbit originally designed for TOPEX/Poseidon. The orbit parameters include repeat period 9.9156 calendar days (unfortunately, often stated as 10 days); inclination 66°; repeat track separation at the equator (316 km); and altitude 1336 km. The radial component of precision orbit determination (POD) is on the order of 2 cm for T/P, and Jason-1 results show POD to a level of 1.5 cm. Although these parameters reflect the fruits of multiple years of trade studies by many individuals,[63] at least one unwanted characteristic remains. The K1 tidal alias is very nearly two cycles per year, thus appearing close to geophysical signals associated with seasonal effects. K1 cannot be ignored, as it is the largest diurnal constituent and is second in magnitude only to the dominant lunar constituent.

The constraint on the exactness of an orbit's repeating ground track is determined to first order by the fine structure in the local geoid expressed at the ocean's surface.[75] For example, cross-track surface slopes (gradients) in the geoid may be as large as 2×10^{-4} near the deeper oceanic trenches. In such an extreme case, a cross-track drift of only 1 km would give rise to a 20-cm change in sea-surface height (SSH). In response, algorithms have been developed that correct SSH data for the effects of the larger cross-track geoid gradients. It also is standard practice to constrain an altimeter's cross-track drift to less than 1 km. Repeat tolerance usually is the condition that motivates active orbit maintenance maneuvers.

Non-Repeat Orbit. The precedent for this is Geosat[71]; its first 18 months were devoted to geodesy for which a non-repeat orbit is optimal. Geodetic missions map gravity anomalies reflected in subtle local tilts of the mean ocean surface. These are static mesoscale features of spatial scales less than about 300 km, determined by the topographic features, composition of the sea bottom, and stationary oceanic currents. Data from Geosat have been used to derive the standard bathymetric charts available for the global oceans.[72,80]

Geosat ERM Orbit. The only other family of dedicated missions is that in the Geosat (1985–1989) Exact Repeat Mission (ERM) orbit, the same orbit used by GFO. This orbit has a period of 17.0505 calendar days (sometimes inappropriately

abbreviated to 17 days), 108° inclination, (consequently 160 km track-to-track spacing at the equator), and 784 km altitude.§ From the Geosat orbit, half of the principal tidal constituents alias into unwanted frequencies (near zero, one, or two cycles per year). In particular, the dominant tidal constituent, the common twice-daily lunar tide, is aliased to 317 days, which is close to the annual cycle.[81] Precision orbit determination is good to only about 7 cm, which is relatively large, due in no small measure to failure of the primary onboard GPS navigation subsystem.

Sun-Synchronous Orbits. Sun-synchronous satellites host European Space Agency (ESA) altimeters on ERS-1, ERS-2, and RA-1 on ENVISAT. All share the same orbit: 35.00 calendar days repeat period; 98.5° inclination; and 781-km mean equatorial altitude. Radial knowledge of these sun-synchronous orbits is good to about 5 cm, based on the Delft model.[82] As sun-synchronous altimeters, the largest solar constituent (twice daily) aliases to zero, and all tidal constituents that are primarily dependent on solar forces alias to frequencies close to zero.

Theoretical Foundations. The following paragraphs provide a summary of the key characteristics of a space-based radar altimeter. Examples are drawn from the design of TOPEX.[83,84]

Pulse-limited Altimeters. Figure 18.10 illustrates the pulse-limited condition.[85] The radius r_P of the area delimited on a quasi-flat surface by a pulse of length τ seconds on the Earth of mean radius R_E seen from a relative altitude of h kilometers is

$$r_P = \sqrt{c\tau h/\alpha_R} \tag{18.16}$$

where $\alpha_R = (R_E + h)/R_E$ is a consequence of the spherical observation geometry. For typical satellite radar altimeters, the pulse-limited footprint over a quasi-flat surface is on the order of two kilometers in diameter. The pulse-limited area A_p is

$$A_P = \pi r_P^2 = (\pi c\tau h)/\alpha_R \tag{18.17}$$

As the pulse continues to impinge and spread over the surface, the resulting pulse-limited annuli all have areas equal to that of the initial pulse-limited footprint. Hence, the received power tends to maintain the level corresponding to the peak of the initial response (Figure 18.10). The pulse-limited areas expand in response to increasing large-scale surface roughness, which in the oceanographic context is expressed as significant wave height (SWH). The height accuracy of a pulse-limited altimeter is much less sensitive to (small) angular pointing errors than is the case for a beam-limited altimeter.

Adaptive Tracking. A satellite-based radar altimeter needs to measure the distance accurately, but only for an essentially planar surface, oriented orthogonally to the radar's line-of-sight. Conservative design suggests that the measurement should be concentrated near the reflection from that surface. Hence, ocean-viewing altimeters have a small range window whose position tracks the delay and strength of the

§ The Geosat ERM orbit was chosen for political as much as for technical reasons. It followed the Seasat orbit (1978) that had established a well-known precedent.[70]

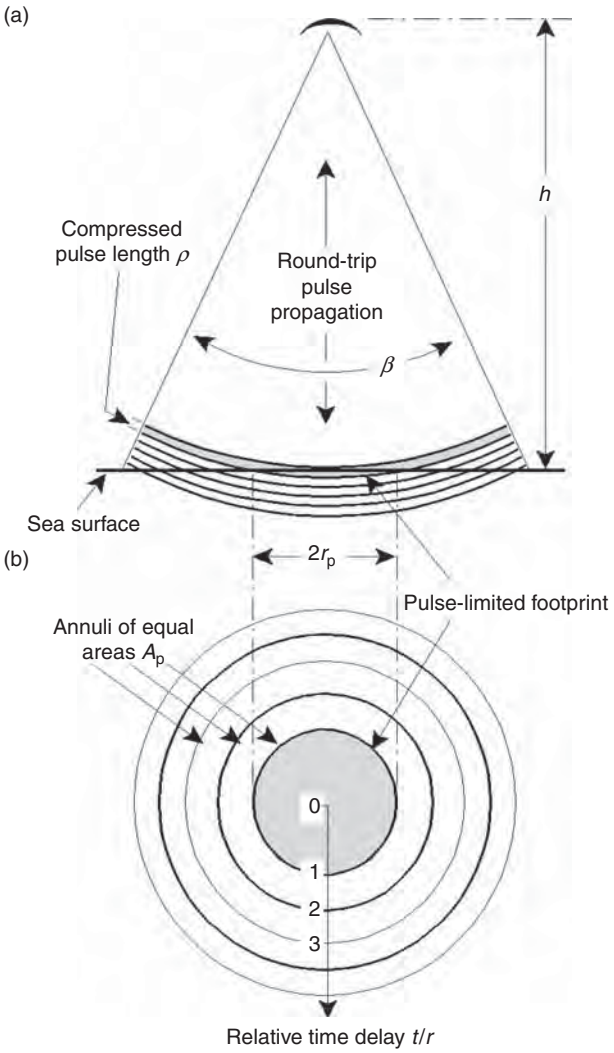


FIGURE 18.10 The pulse-limited condition: Over a nominally level flat surface, the altimeter's short pulse (*a*) reflects first from an area that may be much smaller than the footprint illuminated by the antenna pattern (*b*).

surface reflection.[70] The ocean's surface has a significant wave height of less than 20 m or so. Radar backscatter is primarily specular, typically spanning 3 dB to 20 dB, to cite parameters used in the testing of the TOPEX altimeter. In practice, range-gate delay and backscatter tracking are met with two servo-regulator feedback loops (Figure 18.11). The first loop is a second-order height tracker consisting of range position (alpha tracker) and range rate (beta tracker). The second loop is the receiver gain control (AGC). Altimeter height measurement is given by the setting of the

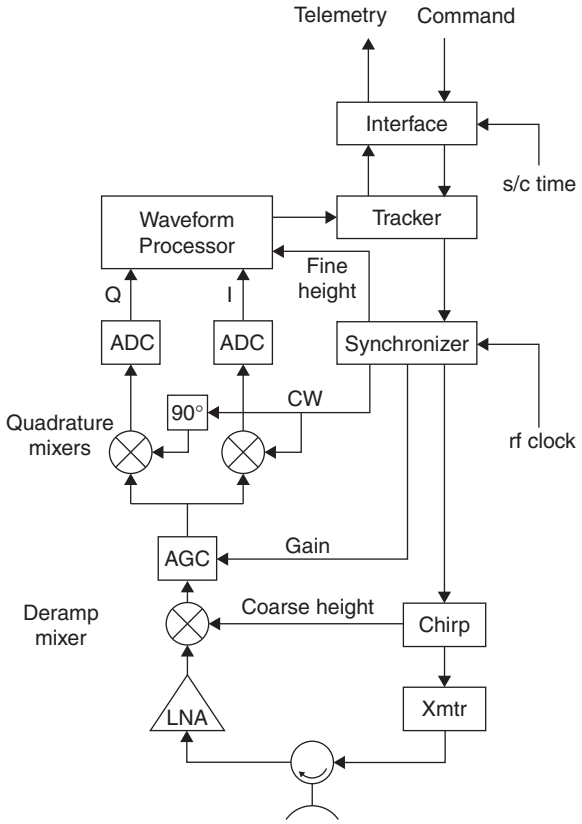


FIGURE 8.11 A generic signal flow diagram for an ocean-viewing radar altimeter. Principal feedback loops include range-gate tracking, both coarse (the alpha loop) and fine range-rate tracking (the beta loop), and mean signal power (the AGC loop).

range delay coarse and fine values, corrected by the remaining height error measured from the waveform's position in the tracker. Surface wind speed and significant wave height[63] are derived from the AGC values and the waveform's shape, respectively.

The precision of an individual height measurement is determined by the combination of range resolution and incoherent waveform averaging. If a single simple short pulse were transmitted, then the height resolution would equal the pulse length. The principal disadvantage of a short pulse is that it contains little energy. The inherent resolution of a pulse is inversely proportional to its bandwidth. Space-based radar altimeters use some form of modulation on the transmitted signal to maintain a large bandwidth within a longer pulse, thus increasing the transmitted energy at no loss of resolution.

Deramp on Receive. Satellite-based radar altimeters present an elegant embodiment of the Stretch technique,[86] which is known as *full deramp* in the field of space-based radar altimetry. This method was first employed by MacArthur in the Seasat altimeter,[70] and has been adopted as the standard technique since then for this type of radar. The distinguishing feature of this technique is a clever trade between the two key

parameters in a large time-bandwidth product (TBP) signal. After reception, demodulation is applied that transforms the "short time, large bandwidth" nature of the original pulse to "long time, small bandwidth" signals. Since the same TBP is maintained, the original resolution is preserved. The method is ideal for altimetry, since the range depth of the ocean's surface is very much smaller than the time available in the pulse repetition period. Clearly, the full deramp technique offers a considerable savings in system bandwidth at all subsequent stages and at no cost in range resolution. The figures cited in Table 18.5 show that the ratio of waveform bandwidth to RF bandwidth for TOPEX is on the order of 0.05%. The TOPEX altimeter design is described more completely in the open literature.[83,84]

Geosat: Geodetic Mission. Radar altimetric data are the basis for state-of-the-art gravimetric variations expressed at the ocean's surface, and consequently, oceanic bathymetry.* The principal objective of a geodetic satellite radar altimeter[73] is to measure the (along-track) slope of the sea surface caused by gravity deflections over spatial scales less than a few hundreds of kilometers (Figure 18.12). These slopes are derived from the SSH measurements summarized above, but this application has its own unique implications for system design. The highlights are reviewed in the following paragraphs.

Sea-surface slope[87] is derived by taking the difference between two neighboring height measurements, where the slope tangent equals "rise over run." The key word for these measurements is *precision*: the standard deviation (noise) of the sea-surface height measurement about its mean value. Height measurement precision is determined by the radar altimeter's post-processing range resolution and by the amount of averaging available for each estimate. Note that a precision measurement may still have poor accuracy, if its mean value is biased away from the correct value. When comparing two neighboring height measurements, any constant bias is cancelled by subtraction as long as the error is the same for both measurements. The sea-surface slope measurement problem is challenging because the desired slope signals are as small as one microradian, equivalent to a 6 mm height differential (rise) for each 6 km along-track separation (run).

In addition to height precision, geodetic altimetry requires smaller along-track resolution than a conventional altimeter and an orbit that accumulates dense cross-track coverage. The altimeter's footprint resolution should be smaller than about 6 km,

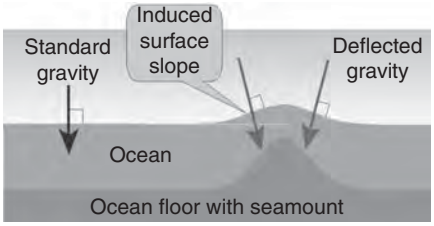


FIGURE 18.12 When averaged and stripped of dynamic (current-driven) features, the mean ocean surface is a direct expression of the local gravity gradient. State-of-the-art radar altimeters can measure the resulting slopes to a 1-microradian precision.

* Literally, measurement of the distance between the mean ocean surface and the local sea floor.

which corresponds to the minimum half-wavelength scale of the observable perturbations in the ocean's mean surface due to spatial variations in the Earth's gravity. The orbit should not repeat for ~1.2 years to yield an average ground track spacing of 6 km, again in respect of the gravity signal at the ocean's surface. The orbit's inclination should be near 50°–63° to resolve north and east slopes nearly equally and to cover the lower latitudes where existing data are inadequate. Note that oceanographic radar altimeter missions (TOPEX/Poseidon, Jason-1, ERS1/2, ENVISAT, and Geosat ERM/GFO) are normally placed into exact-repeat orbits (10 to 35 days), and as a consequence, have widely spaced (80 km to 315 km) ground tracks. Such orbits cannot resolve the short-wavelength two-dimensional surface slopes required for useful geodesy.

Since absolute height accuracy is not required, geodetic radar altimeters can be relatively basic instruments.[88] They do not need to compensate for propagation delays; hence they need only one frequency, and they do not necessarily need a water vapor radiometer (WVR). Indeed, a simple instrument is preferred; it has been shown that efforts to correct for path delays usually add noise to slope estimates.[89] Geodetic measurements provided by the Geosat and ERS-1 (both single-frequency altimeters with no WVR) furnished the best resolution oceanic geodesy and bathymetry available up through at least 2010 for the open ocean. Their resulting bathymetric resolution is limited to about 25 km north-south and poorer resolution of east-west slope components. These results reflect the less-than-optimum resolution, waveform precision, and orbit inclination of those two altimeters. Geodetic resolution at the ocean's surface can be no finer than about 6 km (half a wavelength), a limit that is determined by the average depth of the ocean.

CryoSat: Ice Sheet Mission. Pulse-limited space-based radar altimeters work best over relatively mild topographic relief of mean slope zero, such as the ocean's surface. Over ice sheets or terrestrial surfaces, performance is degraded. Unwanted characteristics include footprint dilation over rougher terrain, height errors in proportion to surface mean slope, and the tendency of the minimum range measurement to hop from one elevated region to another (without the control or knowledge of the data analyst). Beam-limited techniques, of which laser altimeters are extreme examples, circumvent these problems, but may imply their own set of disadvantages.

A major potential application of radar altimetry is to monitor the height of extensive ice sheets, as found in Greenland or Antarctica. Approximately 95% of these land-ice surfaces have slopes less than ~3°, which, although small, is sufficient to trick a conventional altimeter into very large height errors. For example, an unknown 1° slope would lead to a 120-m surface height error, which is unacceptable if cm-level interannual changes are the objective.

The CryoSat altimeter[90] is the first space-based radar altimeter designed to operate over ice (Figure 18.13). Its payload instrument is the SAR/Interferometric Radar ALtimeter (SIRAL), which has three modes: Conventional, SAR, and Interferometric. The Conventional mode (pulse-limited, described in the forgoing paragraphs) reflects its Poseidon heritage. The SAR mode is based on the delay-doppler architecture,[91] which offers advantages in precision, resolution, and along-track surface slope tolerance. The Interferometric mode[92] is designed to measure the cross-track surface slope component. Both of these advanced altimeter modes have been demonstrated with the D2P airborne altimeter.[93]

Unlike previous radar altimeter missions, CryoSat will downlink all altimetric data with essentially no onboard processing. Data from each of the three modes are

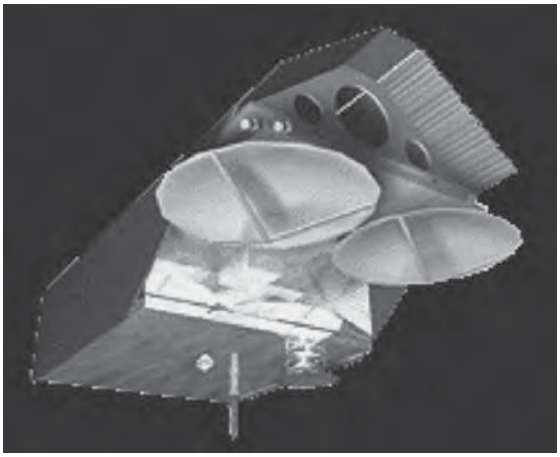


FIGURE 18.13 The CryoSat satellite and its SIRAL altimeter: The two antennas (Interferometric mode) are athwart the velocity vector, so that the differential phase (modulo 2π) of the first return indicates the cross-track position of the minimum-range reflecting surface, which, in effect, is a measure of the cross-track component of the mean surface gradient. (*Courtesy of European Space Agency*)

processed in ground-based facilities, since the more complicated waveforms from icy surfaces require iterative development of suitable processing algorithms by investigators prior to retrieving the desired information. The conventional mode is used for the open ocean (for calibration and sea-surface height reference purposes) and the central continental ice sheets that are relatively level. The interferometric mode is reserved for the more steeply sloping margins of the ice sheets. The synthetic aperture mode is used primarily over sea ice, where its sharper spatial resolution and precision support measuring the difference between the sea level and the top surface of floating ice (freeboard). Because the density of ice is relatively well-known, such freeboard measurements can be inverted to estimate ice thickness.[94]

18.4 PLANETARY RADARS

The history of planetary imaging radars is summarized in Table 18.6. Venus has been the most popular destination,[95] largely because it is cloud-covered (hence its surface is not observable by optical means), its mass and size are similar to those of Earth, and spectral observations have shown that its atmosphere is ~98% CO_2, suggesting that a greenhouse effect could have overwhelmed what might have been a more hospitable planet. A popular theme for planetary exploration by radar is the search for evidence of water (especially water-ice).[96,97] Data from radar missions are available through NASA's Planetary Data System.[98]

Flight Systems. *Venera-8* performed the first simple radar measurements of Venus from a spacecraft. *Venera-8* carried a pulse-modulated radar altimeter, which returned 35 readings during its descent from orbit to its demise on the surface of Venus.

TABLE 18.6 Planetary Radars

Mission	URL	Planet	Year	Radar
Venera -8;9/10 (USSR)	1	Venus	1972, 75	Radar altimetry
Pioneer Venus Orbiter (PVO) (USA)	2	Venus	1978–92	ORAD: Altimeter (also coarse imagery); 17 cm
Venera 15/16 (USSR)	3	Venus	1983–84	SAR and altimeter; 8-cm wavelength
Magellan (USA)	4	Venus	1990–94	SAR: 12.6 cm (125 m, 75 m pixels), 95% coverage
Clementine (USA)	5	Moon	1994	Bistatic scatterometer experiment; 6 cm
Cassini (USA)	6	Titan	2004	TRM: 2 cm, SAR (resolution 0.35–1.7 km) and altimeter
Chandrayaan-1 (India)	7	Moon	2008	Forerunner Mini-RF (USA): 12-cm SAR, scatterometer
Lunar Reconnaissance Orbiter LRO (USA)	8	Moon	2008	Mini-RF: SAR (12 cm and 4 cm), imager and interferometer

1. http://www.mentallandscape.com/V_RadarMapping.htm
2. http://heasarc.nasa.gov/docs/heasarc/missions/pvo.html#instrumentation
3. http://en.wikipedia.org/wiki/Venera_16
4. http://www2.jpl.nasa.gov/magellan/
5. http://filer.case.edu/~sjr16/advanced/20th_close_clementine.html
6. http://saturn.jpl.nasa.gov/spacecraft/instruments-cassini-radar.cfm
7. http://www.lpi.usra.edu/meetings/lpsc2006/pdf/1704.pdf#search=%22chandrayaan-1%20radar%22
8. http://lunar.gsfc.nasa.gov/missions/scandinst.html

The capsule's trajectory was estimated by doppler radio readings and aerodynamic calculations, and by subtracting this from the absolute radar altitude readings, a ground profile could be measured. Readings span a vertical range of 45.5 km down to 0.9 km, during which time the capsule drifted horizontally for a distance of about 60 km. Analysis of the return pulses yielded estimates of elevation variations of the over-flown surface. *Venera-9/10* (1975) demonstrated the first bistatic planetary radar observations. The *Venera-9/10* orbiters mapped 55 strips of the surface of Venus, ranging from 400–1200 km long and 100–200 km wide. A 32-cm wavelength radio wave was beamed at the surface by the telemetry antenna, and both the direct and reflected signals were recorded by Earth-based receivers. The first analysis of these data gave one-dimensional measurements of terrain shape, with a resolution of 20–80 km.

Pioneer Venus was host to 17 experiments (with a total mass of 45 kg), including a radar altimeter (ORAD), which also produced rudimentary surface maps as the radar beam was scanned in the plane orthogonal to the orbit by the spacecraft's 5 RPM spin-stabilization. The radar package required an average 18 W input power and had a mass of 9.7 kg. Peak transmitted power was 20 W. The X band and S band communications system used a despun antenna (~1 m diameter dish). The radar altimeter provided many years of data with a height accuracy of 150 m, which was the best available information on Venus' surface figure until *Magellan*. The altimeter's waveform strength and shape were analyzed to estimate surface electrical conductivity and meter-scale roughness properties.[99] The spacecraft's 24-hour orbit was highly elliptical (~200 km periapsis and ~22,900 km apoapsis[†] for most of the mission). Radar data were collected only below 4700 km altitudes, with a resolved footprint on the surface of 23 km along track and 7 km across track.

[†] *Periapsis* and *apoapsis* are, respectively, the points along an elliptical orbit that are closest to and farthest from the gravitational center of the system.

Venera 15/16, simultaneous "twin" missions, were the first space-based SAR mappers of another planet. They imaged the area from the north pole down to about 30°N latitude over 8 months of operations.[100] Their radars had two modes, imaging and altimetry, operating at 8-cm wavelengths. Imaging resolution was ~1 km. Each 4000 kg spacecraft was a cylinder 5 m in length. The synthetic aperture radar antenna was a 6×1.4 meter parabolic cylinder reflector, powered by an 80-watt traveling wave tube amplifier. Rather than pulse or chirp modulation, *Venera* used a continuous transmission modulated by a coded sequence of 180° phase shifts. The received signal was digitized into 2540 complex numbers (4 bits I, 4 bits Q). A radar look was taken every 0.3 seconds and stored in a RAM memory buffer. To keep up with this data rate, recording alternated between two onboard tape recorders. Data were downlinked each orbit (~9 MBytes) and then processed in ground-based facilities. Each imaging pass generated 3200 radar images that were combined into a survey strip of 120 km wide by 7500 km long. These were subsequently combined into mosaics. The spacecraft included a 1-m diameter parabolic dish antenna for the radar altimeter. After the orbits were accurately determined, the altimeters were switched to high-resolution mode. A 31-element phase modulation was used, with a height ambiguity of 7.15 km. In later phases of processing, doppler frequency analysis narrowed the effective footprint to 10 km by 40 km. The altimeter operations (interleaved with those of the imager) resulted in the first radar altimetric map (of the northern 1/3) of Venus comprised of more than 400,000 individual measurements. The communications system used a dedicated 2.6 m radio dish antenna. The Venera spacecraft were in approximately ~24-hour polar orbits with a periapsis ~1000 km, at 62°N latitude, and apoapsis ~65,000 km.

Magellan mapped over 98% of the surface of Venus (Figure 18.14) with imaging resolution an order of magnitude better than the earlier *Venera* missions.[101]

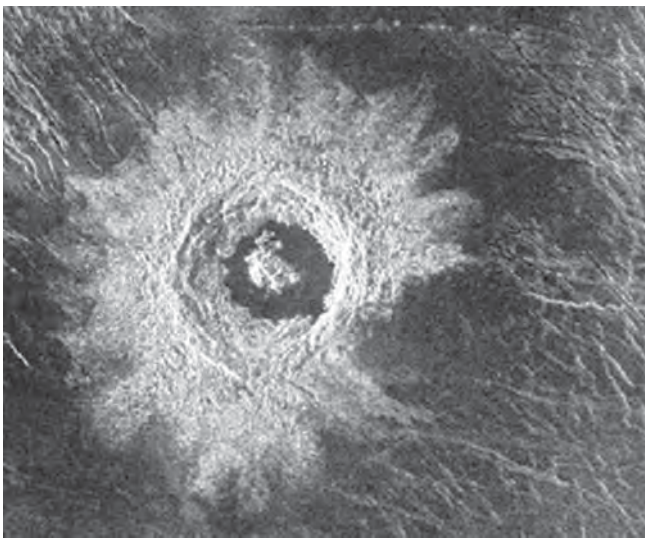


FIGURE 18.14 The impact crater Golubkina on the surface of Venus imaged by *Magellan* (S band, HH-polarized). The 30 km (19 mile) diameter crater is characterized by terraced inner walls and a central peak, typical of large impact craters on Earth, the Moon, and Mars. Rough ejecta give rise to strong radar return, a blessing for those interested in planetary geology. (*Courtesy of NASA*)

Altimetry and radiometry data also measured the surface topography and electrical characteristics. *Magellan*'s elliptical orbit was inclined at 86°, which allowed virtually full access to the surface by the side-looking SAR. By the end of the mission, *Magellan* had returned more data than all prior planetary missions combined. The radar operated in three modes—imager, altimeter, and radiometer—interleaved during each pass.[102] The X band data downlink supported data rates of 268 kbit/s or 115 kbit/s. The 3.7-m diameter high-gain antenna was used for both the radar and for telecommunications. Spacecraft mass was 1035 kg; the radar mass was 335 kg. Input power was 210 W at 28 VDC. *Magellan* operated at S band (2.385 Ghz), radiating a peak power of 325 W. Nominal pulse length was 26.5 μsec, with PRF (4400 Hz–5800 Hz) selectable to accommodate the wide variations in range and incidence necessitated by the elliptical orbit. It achieved 150-m resolution in SAR mapping mode, 30-m height resolution in altimeter mode, and 2°C in radiometer mode. All *Magellan* data are available through the Planetary Data System.[98]

Cassini, a multimode radar mapper[103] leveraged from *Magellan*'s heritage, was included in the 12-instrument payload of the Cassini-Huygens Mission, which was launched in October 1997 and started its four-year tour of Saturn and its moons in July 2004. The motivation for the *Cassini* mapper was the same as that for *Magellan*, namely, to make measurements of the surface of Titan through its dense cloud cover. During its extensive tour of the Saturnian system, the Cassini-Huygens mission was to complete 35 flybys of Titan, of which 29 will be at the closest approach of less than 4000 km, of which 15 will have minimum altitudes of ~1000 km. The first close flyby was in November 2004, from which the first radar images of the surface were collected. The radar antenna uses the 3.66 m H-polarized high-gain telecommunications antenna, a strategy similar to that pioneered on *Magellan*. Seven beams, each at different frequency, width, and boresight orientation, are required to support multiplexed altimetry and scatterometry as well as imaging and radiometry. The radar's mass is ~ 40 kg, and its input power requirement is ~110 W. Peak data rates are on the order of 360 kbit/s. All modes operate at K_u band (13.8 GHz). In its most favorable lower altitude imaging geometry, ground range and azimuth resolutions are on the order of 0.5 km, at 4 looks. At higher altitudes, more looks are gathered to partially offset the degraded resolution. The radar's noise-equivalent σ^0 ranges from ~ −25 dB at low altitude, to ~ −8 dB at 4000 km altitude. In contrast to the scheme used on *Magellan*, smaller angles of incidence and lower bandwidth are used for the higher altitudes. The lower bandwidth helps to reduce the mean noise level, whereas the shallower incident angle helps to maintain range resolution with smaller radiated pulse bandwidth.

Clementine, one of the first "faster better cheaper" missions, had primary objectives including laser altimetry and optical surface mapping of the Moon as well as technology demonstration. The main payload instrumentation was comprised of four optical cameras, including one with a laser altimeter. Major new information about the Moon was provided from *Clementine*'s data collected during its 71 days in lunar orbit.

Clementine is relevant to this chapter because of a unique bistatic radar experiment conducted with the S band (13.19 cm) RF data system. The lunar south pole was illuminated by *Clementine*'s communications antenna (6 watts, circularly polarized), and the reflections from the specular point were tracked by an antenna of the Deep Space Network (DSN) over four passes. The observed reflected signal characteristics were consistent with normal lunar regolith[‡] on three passes, but data from the fourth pass

[‡] Found virtually everywhere on the Moon's surface, *regolith* is a layer of granular rocky material covering solid rock.

seemed to show an enhanced strength that corresponded to the specular point passing across the floor of the crater Shackelton.[104] Such enhanced reflections, especially in the "unexpected" sense of circular polarization, are indicative of volumetric radar response from very cold frozen volatiles such as water-ice,[105] best illustrated by Earth-based radar observations of Jupiter's icy moons. The subsequent claim that water-ice had been discovered at the Moon precipitated wide interest. When independent analyses of the *Clementine* data failed to reproduce the original result,[106,107] considerable controversy was generated as well. Discovery and/or verification of polar ice deposits on the Moon emerged as a major objective of NASA's exploration program,[108] and if proven, would be an essential resource for habitable outposts at the Moon.

Radar Exploration for Planetary Ice. Volumetric ice gives rise to two unusual radar responses. When illuminated by a circularly polarized field, the predominant backscatter from most natural surfaces has the opposite sense of circular polarization. In the case of volumetric ice, however, the backscatter has the same sense of circular polarization. The classical measure of this effect is the circular-polarization ratio (CPR), σ_{SC}/σ_{OC}, or "same sense over opposite sense" circularly polarized backscatter strengths.[97] The total reflected power from volumetric ice is relatively strong, at least for "clean" deposits. Both the polarization and radar brightness effects are explained by the coherent opposition backscatter effect[109] (COBE).

Water-ice was predicted many years ago to have accumulated over some two billion years in the floor of lunar craters or other features whose depth and latitude kept them in permanent solar shadow. The only source of heat for those regions would be background starlight and energy from the Moon's interior, so that the ambient temperature would be no more than ~75 K. As water-ice enters such a cold space from comets, it accumulates. This process generally is accepted as an explanation of the radar-bright response from Mercury's polar craters, for example, as observed by Earth-based radar telescopes. Due to the small axial tilt of the Moon, however, radar telescopes such as Arecibo have no similar opportunity to explore the floors of the lunar polar craters.

The problem is that relatively large CPR is not unique to volumetric ice deposits. Dihedral (two-dimensional) corners also reflect most strongly in the same sense as the incident circular polarization. Naturally occurring dihedrals, such as rough rock formations created by a large impact, could generate false water-ice signatures. To reduce the potential ambiguity of CPR and brightness measurements, the radar observations must be repeatable and should be correlated with other indicators.

Chandrayaan-1 and the Lunar Reconnaissance Orbiter (LRO) include in each of their payloads a "Mini-RF" radar.[110] The version for Chandrayaan-1 is at S-band (12-cm wavelength), with 150-m resolution at 16 looks. The Mini-RF for LRO has two frequencies, S band (12 cm) and X band (4 cm), and two resolutions, 150 m at 16 looks, and 15 m at 8 looks. The LRO radar also includes an interferometric mode, which requires a continuous PRF, in contrast to the burst plan used for the other modes. Both have moderate incident angles (~45 degrees), moderate swath widths (4 km to 8 km), and operate from low altitudes (100 km and 50 km, respectively). Given the applicable small range-velocity product (Table 18.1), the antenna area needs to be only ~ 1 m^2 to satisfy the minimum area constraint (Eq. 18.4). These radars are indeed low mass, at about 8 kg and 12 kg, respectively, including in each case their antenna.

A major objective for both of these radars is to look for evidence of ice deposits in the permanently shadowed areas of the Moon's polar regions. This requires that they must measure the circular polarization ratio (CPR). Hence, they transmit circular polarization, and they are dual-polarized on receive. Their antennas are comprised of passive arrays of H- and V-polarized elements, driven simultaneously 90° out of

phase so that the radiated polarization is circular (either left or right). The received linear polarizations are maintained through the remainder of the system to the image processor output products. This results in the hybrid-polarity architecture[111] that is outlined below.

Magellan: Innovative Venus Mapper. Magellan (Figure 18.15) had to face up to two driving mission restraints: cost and data rate. The first was set by NASA and the U.S. government budget authorities. The second was set by physics, conditional upon the data-rate capabilities of the Deep Space Network.[112] The lesson, of course, is that it is not enough to do good (read "big budget, SBR-based") science, it must be done efficiently and relatively patiently. In the case of *Magellan*, however, these top-level restraints motivated a superb innovative radar design.

Cost. There were several consequences to the severe reduction in approved funds relative to funds requested for the original Venus Orbiting Imaging Radar (VOIR) mission. Rather than a conventional circular orbit, *Magellan* was redesigned for an elliptical orbit (Figure 18.16), which had considerably less associated costs and risk. Also, rather than the original large high-aspect-ratio antenna typical of space-based SARs, *Magellan* was obliged to use a spare circular communications antenna left over from a previous mission. Whereas these may sound like benign if not trivial modifications, they necessitated paradigm-shifting innovations by the SAR design team.

Antenna. The key element in any space-based SAR design is the antenna. It determines range swath coverage and required minimum transmitter power, and has a major influence on resolution and data rate. The Magellan antenna, a 3.7-m diameter dish, was a significant departure from the conventional highly asymmetric rectangular antennas that were then in favor for SARs.

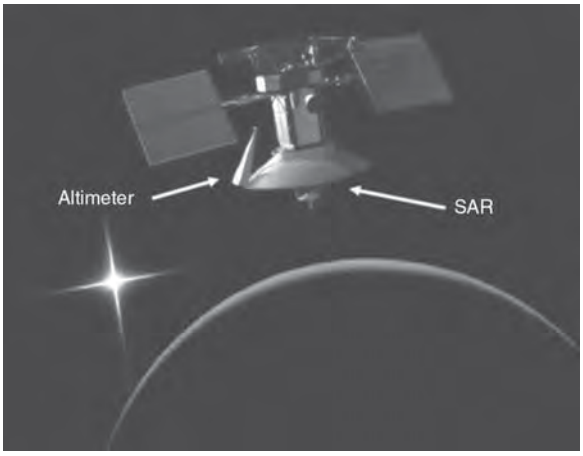


FIGURE 18.15 Artist's impression of Magellan observing the surface of Venus, backlit by the sun, featuring the radar (and high-gain communications) antenna and the smaller horn antenna for the altimeter (*Courtesy of NASA*)

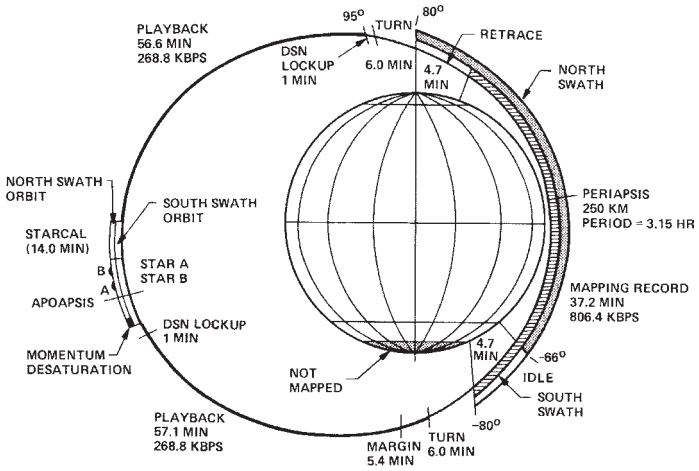


FIGURE 18.16 Details of *Magellan*'s elliptical orbit, dominated by ~37 minutes of data take and 2 hours of data downlink during each 3.3-hour orbit period (*Graphic courtesy of NASA*)

The vertical dimension of the antenna (together with range, incidence, and wavelength) determined the width of the illuminated swath on the surface. This provided an upper bound on the imaged swath width, which, in the case of *Magellan*, was less than the swath actually illuminated. The imaged swath width was chosen to be somewhat wider than the orbit-to-orbit translation of Venus' surface due to planet rotation. The period of the elliptical orbit was selected to be about three hours, so that in succession each imaging pass overlapped and extended the surface imaged on previous passes.

In the along-track direction, the antenna size had far-reaching consequences. The theoretical (single-look) azimuth resolution available from the *Magellan* antenna could be about 1.6 meters (Section 18.2). This is on the order of 75 times smaller than the science requirement of 120 meters azimuth resolution. It followed that single-look 120-m data could be collected by operating the radar only 1/75 of the time. If N_L looks were required, then the implied operating duty factor would be $N_L/75$, larger by a factor of N_L. The result is the burst mode (Section 18.2), in which the radar operates at much less than 100% of the time. Burst mode is a standard strategy for SARs designed to explore the Moon or other planetary bodies, using antennas whose along-track dimension is much smaller than the azimuth resolution required of the data. The looks parameter N_L deserves special comment in the SBR context, summarized below.

The *Magellan* antenna was used as the high-gain antenna for data downlink as well as for the SAR. During the high-altitude segment of each orbit, the spacecraft was oriented to point the antenna toward Earth. The antenna had two feeds, one at S band (HH-polarization for the radar) and X band (circular-polarization for telemetry). After each pole-to-pole data collection, the telemetry transferred the accumulated data to one of the three Deep Space Network (DSN) receiving stations, so that, in consequence, the data dump kept pace with the data take, per orbit.

Orbit. For reasons directly attributable to principles of orbital mechanics, it was much less costly to go into an elliptical orbit at Venus rather than a circular one. The fact that it would be considerably more difficult to get decent imagery successfully

from an elliptical orbit was of little concern to the budget-masters, although the elliptical orbit constraint attracted more than a little attention from the SAR design team. The *Magellan* radar had to adapt to variations in relative altitude from approximately 250 km (near the equator) to more than 2000 km (over the poles). Quite reasonably one might expect that the resulting imagery—let alone the implied timing and scaling issues—would suffer as a consequence. Thanks to the mission design, however, *Magellan* image quality is surprisingly consistent, pole to pole. The reasons provide an important object lesson.

Since the Magellan radar was operated in burst mode, it was convenient (as well as necessary) prior to each burst to set the mode parameters, which, in general, varied from burst to burst. Critical parameters included PRF, range gate, burst length, burst period, and spacecraft roll. The parameter files were prepared in advance, based on the data collection geometry, and turned into commands to be generated by the radar mapping sequencing software. The result was a set of about 1000 unique configurations, each tied to specific segments of the orbit. In operation, suitable commands were preloaded into the onboard SAR control processor for each three-day mapping interval.

In order to offset the tenfold altitude change on SAR image quality, the SAR operating profile was designed to exploit a variety of incident angles. At the start of each imaging pass at high altitude, the incidence was slewed from steep near the pole, to shallow in the neighborhood of the equator, back to steep when approaching the opposite pole. This incidence variation helped to offset the large change in radar range to the imaged swath, but it meant also that the range resolution as expressed on the surface varied as a function of latitude. For a constant radar bandwidth, the effective surface range resolution at shallower incidence is improved over that achieved at steeper incidence. Fortunately, at steeper incidence, and hence at longer ranges, there was more time available to gather more looks.

Image Quality. Looks and resolution work together to determine the image quality (read "geophysical information potential") of SAR imagery over natural terrain composed of distributed scatterers. The governing expression is the SAR image quality parameter[31,113]

$$Q_{\mathrm{SAR}} = \frac{N_L}{r_{\mathrm{Rg}} r_{\mathrm{Az}}} \tag{18.18}$$

where N_L is the number of (statistically independent) looks, and r_{Rg} and r_{Az} are the range and azimuth resolution, respectively, on the surface. Here the important generalizable lesson is that an increase in number of looks can be applied to offset a decrease in range resolution (within reason and in this kind of exploratory space-based SAR data). Note that both looks and resolution require support in bandwidth. It follows that Q_{SAR} is proportional to the product of the range and azimuth bandwidths, hence proportional to the (two-dimensional) information capacity of the radar in the Shannon sense. This principle was applied with great profit to the *Magellan* SAR design,[114] as illustrated by Table 18.7. From that table, one can verify that the image quality of *Magellan* data varied by no more than ~±2% pole to pole, in spite of large variations in radar range, incident angle, and ground range resolution.

Data Rate. *Magellan* SAR data were relayed from Venus to Earth via the DSN. This major communication system imposed a working limit of about 270 kbit/s on the SAR data telemetry. Whereas this may seem rather large, it is minuscule by

TABLE 18.7 Image Quality: Magellan Resolution and Looks

Altitude (km)	Incidence (degrees)	r_R (m)	r_A (m)	Looks NL	$Q_{\text{SAR}}\, N_L/(r_R\, r_A)$
250	52	110	122	4.8	3.6×10^{-4}
500	39	137	121	6.0	3.6×10^{-4}
1000	28	181	121	8.5	3.9×10^{-4}
1750	21	247	120	11.6	3.9×10^{-4}
2100	19	270	120	13.5	4.0×10^{-4}

space-based SAR standards. For example, the digital data rate for the ERS-1/2 and RADARSAT Earth-observing SARs is on the order of 120 Mbit/s. Newer designs claim up to 400 Mbit/s. Since the *Magellan* objective was to image a substantial portion of the surface of the entire planet at 120-m resolution within the length of the primary mission, the DSN data-rate capability emerged as the toughest requirement on the entire system.

Raw SAR data rate is proportional to the image quality factor, swath width, spacecraft velocity, and the number of digital bits per data sample. Of course, mean data rate can be relaxed if the data from any given pass can be collected and then played back at a slower rate. Although this strategy was exploited by the *Magellan* design, tolerance for which was set up by the elliptical orbit, it was not sufficient to solve the limited data-rate problem.

The only degree of freedom remaining in the data-rate budget was the number of bits retained for each sample in the raw SAR datastream. Working back through the DSN constraint, it turned out that there were only two bits available for *Magellan*'s raw SAR data.

Yes—a 2-bit SAR! And again, the unique characteristics of an orbital SAR made this an acceptable solution for the *Magellan* design. The key requirements for this paradigm are two: (1) an effective signal encoding that is very large and (2) an imaging environment that is dense and dominated by distributed random backscatter. Planetary SARs (and many Earth-observing space-based radars) easily satisfy these requirements. One measure of an imaging radar's extensive signal encoding is the product of its range and azimuth time-bandwidth products, or equivalently, its potential two-dimensional compression ratio. In the case of SAR data, this ratio is given by the area of the instantaneous pulse footprint (antenna width by projected pulse length) divided by the resolved cell area (the product of azimuth and range resolution). This ratio varied by mode for *Magellan*, but, in general, was well in excess of 10,000, which turns into a gain in dynamic range from the signal domain to the image domain of 40 dB.

The "2-bit method" was based on a Block-Adaptive Quantizer[115] (BAQ). This in effect implemented an automatic gain control (AGC) that selected the most influential digital samples (bits) from the raw SAR datastream. As designed for *Magellan*, the data were digitized into 8 bits, in-phase and quadrature (I&Q). The analog-to-digital stage was followed by the BAQ operation, which selected the two most significant bits in each (I&Q) data pair, relative to a mean signal level that had been established from the previous burst of received data. The mean signal level varied very slowly burst to burst since adjacent bursts covered essentially the same scene elements. The mean signal level (AGC setting) was included in the header for each burst's data record to be used in subsequent SAR image formation.

Although the dynamic range of the raw SAR data out of a BAQ operator is severely limited, the potential dynamic range of the resulting image data is much larger: it is bounded above by the product of the input dynamic range and the two-dimensional compression ratio of the SAR data. Thus, the dynamic range capacity after processing for *Magellan* imagery was in excess of 40 dB. This was well illustrated in the many thousands of image frames formed from *Magellan*'s SAR data.

Hybrid-polarity Architecture. A leading high-level objective of a radar destined for the Moon, Mars, or any other planetary body is to maximize its measurement potential, while also minimizing its resource demands (principally power and mass). If sensitivity to frozen volatiles is a requirement levied on a planetary exploratory radar, then the system should be dual-polarized and must transmit circular polarization. As reviewed in Section 18.2, a dual-polarized radar maximizes its measurement capabilities only if it retains the relative phase as well as the magnitudes of the two received amplitudes, such as E_H and E_V in the linear polarization basis. It has been known since 1852 that a quasi-monochromatic EM field can be fully characterized by the four Stokes parameters.[116] In terms of linearly polarized received data, the Stokes parameters are

$$\begin{aligned} S_1 &= \left\langle |E_H|^2 + |E_V|^2 \right\rangle \\ S_2 &= \left\langle |E_H|^2 - |E_V|^2 \right\rangle \\ S_3 &= 2\,\mathrm{Re}\left\langle E_H E_V^* \right\rangle \\ S_4 &= -2\,\mathrm{Im}\left\langle E_H E_V^* \right\rangle \end{aligned} \tag{18.19}$$

where * denotes complex conjugate, and the carats < > indicate an average taken over several samples. Clearly, the relative phase between the two polarizations is an essential factor for two of the four Stokes parameters. Data expressed through the Stokes parameters are well-suited to being exploited by matrix decomposition methodology.[60,117]

For a given transmitted polarization, the values of these Stokes parameters are invariant with respect to the polarization basis of the receiver. It follows that the optimum architecture is hybrid-polarity[118]: circularly polarized on transmit and linearly dual-polarized on receive (Figure 18.17). This architecture requires less mass and offers greater efficiency than alternatives, while capturing all of the potential information in the backscattered field.

An antenna comprised of two linear arrays (such as H and V) will radiate a circularly polarized field if the sets of elements are driven simultaneously and 90° out of phase, as shown in Figure 18.17. In practice, the amplitude weighting and relative phasing of the arrays will seldom be perfect. As a result, the radiated field will be somewhat elliptical, rather than purely circular. The hybrid-polarity architecture is self-calibrating and, therefore, relatively robust in response to such imperfections. In brief, under the condition that $\sigma^0_{HH} = \sigma^0_{VV}$, the mean signal level in the two receive channels should be equal. (In terms of the Stokes parameters, $S_2 = 0$). The H and V backscatter coefficients will always be equivalent when the radar's illumination is perpendicular to the surface. Hence, any hybrid-polarity radar can set up this condition by the simple expedient of looking down on a horizontal surface during a calibration exercise. Discrepancies in either amplitude or phase will be evident from such data collected over a random distributed scene. There is no need for a known point-target reference in the field of view. The Stokes parameters that correspond to such measurements are sufficient to characterize the relative H/V phase of the transmitted field, as well as that of the receivers.

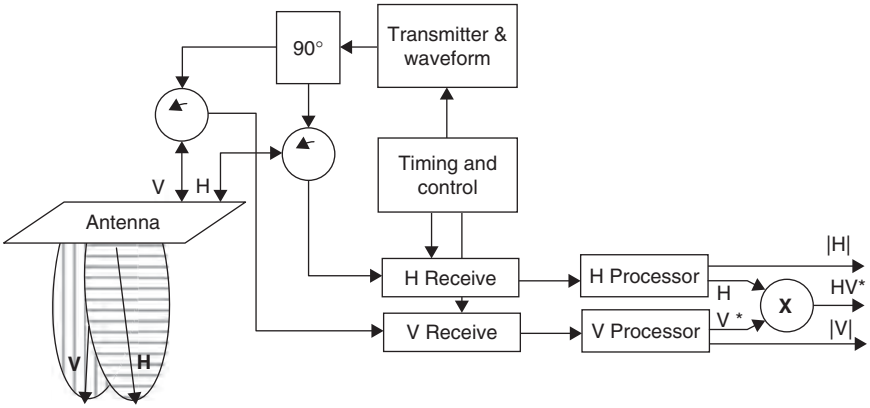


FIGURE 18.17 The hybrid-polarity SAR architecture first used by the Mini-RF lunar radars on *Chandrayaan-1 and LRO*

18.5 SCATTEROMETERS

Space-based remote sensing scatterometers measure the normalized backscatter with sufficient precision and accuracy to deduce the value of one or more parameters of geophysical significance. For example, the power reflected from the ocean back to a radar is a function of surface roughness at the scale of the radar's wavelength, which, in turn, is a function of the local wind.[119] Estimation of wind speed and direction over the open ocean is the most common application for these instruments. A wind scatterometer was adopted by EUMETSAT in the early 2000s as a required operational capability, with operational accuracies of ±2 m/s in wind speed and ±20° in direction. In addition to open ocean observations, calibrated data from this class of SBR have been applied to a variety of large-area surface features, such as determination of sea ice coverage, mapping the boundaries between the principal ice zones of Greenland, or global estimation of tropical deforestation. In all such applications, the emphasis is on measurement of mean reflectivity over large areas, rather than mapping fine spatial detail. These radars typically have resolutions on the scale of 10s of kilometers, supported over swaths of 1000 kilometers or more.

In the ocean application, relatively small changes in radar backscatter may correspond to substantial differences in the retrieved wind information.[§120,121] It follows that the dominant requirement for this class of radar is the accuracy and precision of the received power measurement.[*] However, the next step, vector wind retrievals—transforming the radar backscattered power into accurate estimates of wind speed and direction—is far from trivial. Indeed, the technique fails for the very low wind speeds that do not generate wavelength-scale surface roughness. In the limit,

[§] http://www.eumetsat.int/groups/ops/documents/document/pdf_tm03_rev-scatterometer-w.pdf

[*] Note that both accuracy and precision are required. Radar backscatter by its very nature is a quantity having a large standard deviation that can be reduced only by extensive averaging. Because the geophysical interpretation of scatterometric data often depends on distinguishing between two similar values of σ^0, the results depend critically on reducing the uncertainty in the estimated value as well as getting the average value right.

an oblique-viewing radar such as a scatterometer will generate virtually no backscatter from the sea surface in the absence of wind-driven waves, even if there is a substantial swell in the region.

Vector Wind Retrieval. The normalized backscatter coefficient σ^0 from a wind-roughened oceanic surface depends foremost on the radar's wavelength, the local angle of incidence, and the polarization, respectively. The wind parameters to be estimated are its speed, the relative angle in the horizontal plane between the wind direction, and the radar line-of-sight. The ocean's reflectivity is also a function of other factors, including surfactants (such as oil slicks, either natural or anthropogenic), the air-sea temperature difference, or the presence of large waves such as oceanic swell, but these are of less significance for the present discussion.

The reflection coefficient σ^0 is nonlinear with respect to the wind parameters. Figure 18.18 illustrates the response for one polarization, where the horizontal axis is the relative wind direction, and the vertical axis is the normalized radar cross section. The several curves each correspond to a specific wind speed. In general, upwind and downwind aspects provide stronger backscatter than cross-wind, and the upwind aspect usually is a bit stronger than downwind. Wind vector data have been collected by airborne scatterometers flying in circles (literally!) over instrumented test sites.[122] There have been many attempts over the years to converge on a suitable mathematical model for this behavior, with reasonable success.[123] There are several vector wind retrieval methods in current use, including CMOD-4 and a neural network model.

The data clearly show that measurement of radar backscatter at only one aspect angle is not sufficient to determine wind speed and direction. The Seasat scatterometer used two look angles (separated by 90°), but the resulting retrievals had as many as

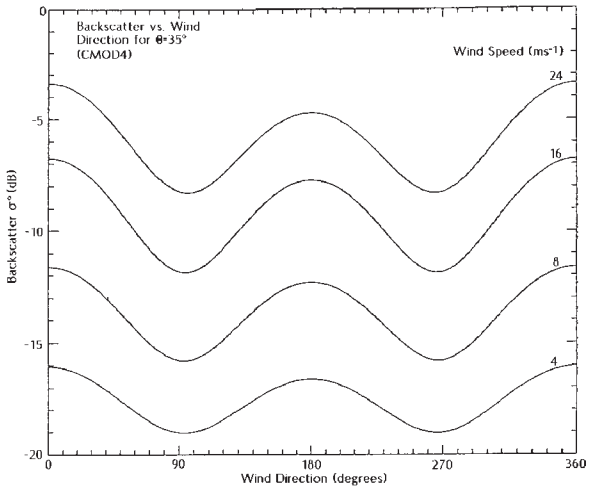


FIGURE 18.18 Typical backscatter strength (vertical axis) from a wind-driven sea surface, as a function of wind speed (modeled data) and the wind direction relative to the radar's look direction (horizontal axis). Similar families of curves correspond to the radar's polarization (usually HH or VV) and angle of incidence. Curves such as these are derived from a model (such as CMOD4) subsequently verified by extensive airborne measurements.

four potential solutions, which presented as directional ambiguities.[124] Space-based wind scatterometers are characterized by differing viewing geometries, each motivated by the need to suppress directional ambiguities, within the constraint of a physically (or financially) feasible implementation.

Measurement Precision. Clearly, σ^0 must be measured with an accuracy and precision of less than 1 dB if useful wind retrievals are to be derived. Accuracy depends on the radar's stability and its calibration. The challenge for space-borne scatterometry is to design the radar such that the precision—the normalized standard deviation of the σ^0 measurements—is sufficiently small. In the standard terminology of scatterometry, the classical parameter is K_P, the normalized standard deviation of the measurement.[125] In the case of a fan-beam scatterometer that employs doppler filtering, the governing expression is

$$K_P = \frac{(1 + 2/SNR + 1/SNR)}{\sqrt{NTB}} \tag{18.20}$$

where N is the number of statistically independent pulses summed into each σ^0 measurement, T is the transmitted pulse length, B is the doppler bandwidth of the measurement cell, and SNR is the signal-to-noise ratio. K_P usually is cited as a percent, where values of 5% or better are the objective. Note for high SNR that K_P converges to $1/\sqrt{NTB}$. At lower wind speeds K_P depends on SNR as well as on the number of statistically independent looks. The details of any expression such as this depend on the underlying statistical model, which is gaussian in this classic expression. However, the general principle is that the scatterometer must provide many independent looks to reduce the standard deviation of the σ^0 measurement, regardless of the statistical distribution of the ocean's backscatter.

Trends. Vector wind data have been adopted by operational meteorological agencies such as EUMETSAT. It is likely that space-based microwave assets will continue to be used for the foreseeable future to provide these data. The success of SeaWinds (discussed later in this section) suggests that the conical scan paradigm will be the basis for future vector wind scatterometer design.

That said, there are two space-based alternatives for measuring vector winds by microwave means, active and passive. Radars often are unpopular on spacecraft that host other instruments, of which some may be compromised by radio frequency interference (RFI) generated by a nearby microwave transmitter. Radars also require more power from their host spacecraft than passive systems. These and related considerations motivated the development of WindSat, a passive instrument that estimates near-surface vector winds by Stokes parameter analysis of the microwave emissivity from the ocean surface.[126] WindSat is one of two instruments aboard the *Coriolis* satellite, launched in 2003. The operational reliability of passive-vector wind data relative to scatterometric measurements remains an open issue.

Scatterometric data are being exploited for many purposes other than oceanic vector winds. Although the "images" generated by a space-based scatterometer may have only 50-km resolution, their wide swath and frequent revisit intervals are well suited to synoptic coverage of global-scale phenomena. The multi-year history of scatterometric radars provides an important data set for climate change studies as well as monitoring seasonal variations. Suitable applications include sea ice cover, large icebergs, continental ice sheets, vegetation, and soil moisture.[127–130]

Flight Systems. *RadScat*[131] (Table 18.8) was the name given to the radiometer/scatterometer portion of the S-193 K_u band instrument aboard Skylab. The top-level objectives of this experiment were (1) to provide the near-simultaneous measurement of microwave backscatter and emissivity of land and ocean on a global scale and (2) to provide engineering data for use in designing space radar altimeters. The equipment shared a common gimbaled antenna. The scatterometer measured the normalized backscatter coefficient of ocean and terrain as a function of incidence ranging from 0° to 48°. Although only sparse coverage of selected sites was possible, the data were sufficient to demonstrate the potential of space-based radar measurement of surface wind vectors over the ocean. The S-193 zone of access was 48º forward and 48º to either side of the spacecraft ground track. For selected measurements, the beam was pointed in the along-track direction to fixed angles of 0º, 15.6º, 29.4º, 40.1º, and 48º, with sufficient dwell time at each angle to permit averaging to achieve approximately 5% precision. RadScat data collected over the Amazon rain forest suggested that the uniformity of the observed backscatter would be a stable calibration reference for space-based radars, which has since been validated as a standard technique.[132]

SASS, the Seasat-A Satellite Scatterometer,[133,134] was the first space-based radar designed specifically to measure oceanic winds. It was a multi-fan-beam instrument, comprised of two sets of two dual-polarized antennas (HH and VV), each ~ 3-m in length whose fan-shaped beams pointed to 45° and 135° to either side of the orbit plane. Incidence spanned 25° to 55°, covering a 500-km wide swath on each side. Calibration data taken over the rain forest were used to reduce antenna gain uncertainty to less than 0.4 dB. Since Seasat was not yaw-steered, the fore and aft footprints were misregistered due to Earth rotation, which reduced the useful swath at lower latitudes to about 400 km. The value of K_p varied from 1% to 3% over the ocean under moderate to high sea states, but degraded to 15% for lower wind speeds and to 50% for very low backscatter from non-oceanic surfaces. The accuracies of wind speed and direction over 4–16 m/s were ±2 m/s and ±20º, respectively. The data were not sufficient to avoid directional ambiguities, however. Nominal resolution was 50 km, determined by the intersection of the antenna pattern and iso-doppler contours. The radar operated at 14.6 GHz, radiating a peak power of 100 W at 17% duty factor from a traveling wave tube amplifier (TWTA). The waveform was modulated CW. The receiver front end was a tunnel diode amplifier, which maintained the noise figure to be less than 5.7 dB at all operating temperatures. Mean input power was 136 W dc; the instrument mass was 102 kg.

TABLE 18.8 Scatterometers

Name	Spacecraft	Country	Year	Antenna	Band	Polarization
RadScat	Skylab	USA	1973, 1974	Pencil beam	Ku	VV, HH
SASS	Seasat	USA	1978	Fanbeams	Ku	VV, HH
ESCAT	ERS-1	Europe	1992-6	Fanbeams	C	VV
ESCAT	ERS-2	Europe	1995-	Fanbeams	C	VV
NSCAT	ADEOS I	US/Japan	1996–1997	Fanbeams	Ku	VV, HH
SeaWinds	QuikSCAT	US	1999–	Conical scan	Ku	VV, HH
SeaWinds	ADEOS II	US/Japan	2002–	Conical scan	Ku	VV, HH
CNSCAT	SZ-4	China	2003	2 conical scan	Ku	VV, HH
ASCAT	MetOp-1	Europe	2006	Fanbeams	C	VV
Scat	Aquarius	USA	2009	3-beam push	L	VV, VH, HV, HH

WS aboard the ERS-1 and ERS-2 spacecraft of ESA denotes the scatterometer mode embedded in their C band AMI radar instrumentation.[14] These scatterometers used three fan-beam antennas, whose footprints were oriented at 45°, 90°, and 135° with respect to the satellite track. The two outer antennas are 3.6-m long. Note that the spacecraft had to be yaw-steered to maintain this footprint geometry over the rotating Earth. The C band operating frequency, although a departure from space-based precedent, resulted from use of the same RF hardware as the SAR mode. As a consequence, simultaneous SAR and scatterometer operations were not possible. In contrast to K_u band, the C band data were less degraded by rain and proved to be more reliable for higher wind speeds.[119] The antenna swaths spanned 500 km, from which the averaged data from the three-look directions were colocated into 50-km resolution cells, posted on a 25-km grid.

ASCAT, the Advanced Scatterometer[135] aboard MetOp-1, is essentially an improved version of the ERS-1/2 instruments. It is a stand-alone radar designed for operational use; it does not have to time-share onboard electronics as was the case with the scatterometer in the AMI suite. It covers swaths on both sides of the satellite ground track. The near edge of these 500-km swaths are offset 384 km from nadir, spanning 25°—65° incidence. The 0.57 dB radiometrically accurate data are averaged to achieve K_p from 3% (high up-wind speed) to 10% (low speed, cross-wind aspect). The resulting vectors for near-surface winds cover 4–24 m/s with accuracy ±2 m/s and ±20°. The radar operates at 5.255 GHz, radiating 10 ms linear frequency-modulated (LFM) pulses at 120 W peak power from combined GaAs FET devices. Only one antenna is active (for 0.2 s) at a time; the operation cycles around the six antennas in sequence. The instrument mass is 270 kg; input power required is 251 W. Thanks to onboard processing, the inherent data rate is reduced from 1.4 Mbit/s to an average of 60 kbit/s to the MetOp-1 payload data-handling system.

NSCAT, the NASA Scatterometer,[134] provided to Japan as part of the Advanced Earth Observing Satellite (ADEOS) payload, was an upgraded version of SASS. NSCAT used six dual-polarized stick antennas (3 m long), four of which were aimed at ±45° (H polarized) and ±135° (Vpolarized) from the surface track, and the two midbeam antenna patterns were aligned at 65° and 115° (each H and V polarized). The third beam on each side helped to remove the fourfold directional ambiguities that plagued SASS wind retrievals. Two 600-km wide swaths were covered, at 25-km resolution. In order to support 25-km along-track resolution, the radar's sequencer had to cycle through all antenna patterns within 3.74 s, resulting in a maximum dwell time of 470 ms within each of the eight footprints. Cross-beam resolution was determined by doppler analysis. However, since the mean doppler offset was a function of antenna orientation (as well as incidence), the return from each direction needed its own LO offset. The antennas' peak gain was 34 dB, directed toward maximum range. NSCAT's mass was 280 kg; required input power was 275 W. The RF system was built around redundant TWTAs, transmitting modulated 5-ms pulses at a PRF of 62 Hz, peak power 110 W.

SeaWinds marks a significant departure from the "stick antennas" of prior wind scatterometers, relying instead on a dish antenna[137] rotating at 18 RPM to sweep its two beams over a nadir-centered swath 1,800 km wide (Figure 18.19). The first SeaWinds mission was aboard QuikSCAT, mobilized by NASA and launched in June 1999 as a rapid response to the premature loss of ADEOS in June 1997. The second SeaWinds was embarked on Japan's ADEOS-II. Both instruments operate at K_u band (13.4 kHz), radiating 110 W, 1.5 ms pulses at 190 Hz PRF, split equally between the two antenna beams. The transmitter is a TWTA, based on NSCAT heritage. The modulated pulse bandwidth is 40 kHz, which is maintained within an 80 kHz filter in the receiver. The received data must be compensated to offset the doppler shift, which varies

FIGURE 18.19 The Seawinds scatterometer, as embarked on Japan's ADEOS-II spacecraft. The radar drives the conically scanned reflector antenna, which takes up most of the real estate on the Earth-viewing face of the satellite.

sinusoidally over 1 MHz during each antenna rotation. Instrument mass is 191 kg; required input power is 217 W. These numbers illustrate the principal advantages of the architecture, providing greater coverage with less mass and power, in contrast to the stick-antenna architecture.[138] System sensitivity accommodates σ^0 in the range –37 dB to –2 dB. The antenna is a 1-m reflector (gain ~40 dB), with two feeds, resulting in a pair of pencil beams at 46° (H polarization) and 54° incidence (V polarization). The illumination geometry also is advantageous because the incidence is the same for all aspect angles. The beam-limited footprints are approximately 30 km by 40 km. Following onboard processing, the average data rate is 40 kbits/s. SeaWinds performance[139] is at least comparable to other wind scats, wind speed and direction accuracy being 3 m/s to 20 m/s at 2 m/s and 20°, respectively. Nominal surface resolution is 50 km; advanced processing[140] reduces this to ~25 km. Although the swath is 1800 km wide, the variety of aspect and polarization coverage limits the science-compliant wind vector retrievals to strips from 250 km to 800 km either side of the ground track.

Aquarius. The *Aquarius* mission is designed to map sea surface salinity, for which L band radiometric sensing of the ocean's emissivity is the primary measurement. However, emissivity is a function of the roughness of the surface,[141] as well as its temperature and dielectric constant, which is the variable of interest. The Aquarius payload includes an L band (1260 MHz) scatterometer to measure surface roughness. The scatterometer and radiometer share the same 2.5-m diameter reflector illuminated by three offset feeds, which generate three side-looking beams (at 29°, 38°, and 45°, incidence), thus sweeping out strips of coverage as the spacecraft moves along its orbit. The scatterometer is fully polarized (HH, HV, VH, and VV). Peak transmitted power is ~250 W, pulse length is 1 ms, sufficient to support sea surface σ^0 over the range 0 dB to –40 dB. Resolution is modest, at ~150 km. The scatterometer and radiometer share electronics. Combined instrument mass is ~400 kg, and required prime power is ~450 W.

18.6 RADAR SOUNDERS

In its most general form, a radar sounder is a device whose transmissions are designed to penetrate the volume of a target medium, from which the waveform of the resulting backscatter indicates variations in dielectric contrasts as a function of depth.† As a sounder passes over an illuminated region, the sequence of ranging waveforms generates a profile, which is a two-dimensional reflectivity cross section of the surveyed volume. Penetration depth in general increases with wavelength and also with radiated power. On the other hand, reflectivity depends on the dielectric contrasts between internal layers; a material's dielectric constant is also a function of wavelength. It follows that space-based radar sounders must choose a frequency and bandwidth that balance the often conflicting requirements of penetration, reflectivity, and resolution, under the constraints of available power and antenna aperture.

The space-based radar sounders highlighted in Table 18.9 fall naturally into two groups: subsurface and atmospheric/ionospheric. It is evident that the subsurface sounders all are at relatively low frequency, in contrast to the atmospheric sounders at much higher frequency. The ionospheric sounding mode of MARSIS is a special case, elaborated below.

Subsurface Sounding–flight Systems. Subsurface sounding from a space-based radar is considerably more challenging than from a surface-mounted GPR (Chapter 21) or a very low altitude aircraft. Of course, once radar signals penetrate the surface, the usual volumetric attenuations and reflections occur. As with any GPR, large dynamic range is required, since the signals of interest may be weaker than competing returns by 50 dB or more. Radar sounding to appreciable depth is possible only in dry materials such as lunar regolith or in very cold low-loss ice.

TABLE 18.9 Radar Sounders

Instrument	URL	Spacecraft	Years	Objective	Frequency
ALSE	1	*Apollo-17*	1972	Lunar subsurface	5, 15, 150 MHz
MARSIS	2	*Mars Express*	2003–	Mars subsurface	1.8, 3, 4, 5 MHz
SHARAD	3	*MRO*	2005–	Mars subsurface	15 - 25 MHz
LRS	4	*SELENE*	2007	Lunar subsurface	5 MHz
PR	5	*TRMM*	1997–	Rain	13.8 GHz
MARSIS	6	*Mars Express*	2003–	Mars ionosphere	0.1–5.4 MHz
CPR	7	*CloudSat*	2006	Earth cloud profiles	94 GHz
DPR	8	*GPM*	–	Rain	13.6, 35.5 GHz

1. http://nssdc.gsfc.nasa.goc/database/MasterCatalog?sc=1972-096A&ex=4
2. http://sci.esa.int/science-e/www/area/index.cfm?fareaid=9
3. http://mars.jpl.nasa.gov/mro/overview/
4. http://www.jsfws.info/selene_sympo/en/text/overview.html
5. http://trmm.gsfc.nasa.gov
6. http://sci.esa.int/science-e/www/area/index.cfm?fareaid=9
7. http://cloudsat.atmos.colostate.edu/
8. http://gpm.gsfc.nasa.gov/dpr.html

† The term *sounder* usually associated with acoustic echo sounding, derives from many centuries of oceanic depth measurements using lead lines and the like. The logical extension from acoustic to electromagnetic methodology is a small step in comparison.

In addition to the usual GPR considerations, two space-specific issues arise, neither of which can be solved by the usual expedient of increasing the radar's transmitted power. The first problem is clutter. From orbital altitude, scatterers on the surface far from nadir may generate strong backscatter that appears at the same radar range as the signals reflected from depth, as suggested in Figure 18.20. The problem is compounded with increasing altitude and by the fact that at the long wavelengths required for deeper penetration, the size constraints on a spacecraft-mounted antenna dictate that the illumination pattern will have little or no directivity. Processing over groups of returns to reduce the effective width of the beam can be helpful. Doppler-based techniques are applicable to the along-track direction.[142] However, the orbit's altitude-velocity parameter (Table 18.1) may be so large that the available ambiguity-free range-doppler window is too small to be useful. Narrowing the effective antenna pattern in the cross-track direction presents even more of a challenge.

The second problem is range sidelobes. From orbital altitudes and from a realistic spacecraft, it is not feasible to transmit a simple short pulse that has sufficient energy to generate useful reflections from depth. Long modulated pulses are the only practical method. Unfortunately, the specular component of surface reflection usually is very strong at the long wavelengths required for subsurface sounding. Pulse compression of the surface return generates range sidelobes which appear at depths that could easily overwhelm the weaker reflections from the internal structure. The standard strategy to mitigate this problem is rigorous sidelobe control, requiring severe pulse amplitude weighting and strict control of phase and amplitude linearity.

Due to the large footprint of a space-based sounder, and the need to salvage all possible contributions of signals from depth, it is standard practice to assume that the dominant in situ returns arise from specular scattering, hence from extended horizontal layers.[143] The contributing area is determined by the radius r_F of the first Fresnel zone, $r_F = \sqrt{h\lambda/2}$ in free space, where h is the altitude of the radar above the surface. The radius is somewhat larger within the medium, since the spherical wavefront is flattened due to the slower speed of in situ propagation.

ALSE. The Apollo Lunar Sounder Experiment[144,145] was a combination imager and sounder, operating at wavelengths of 60 m, 20 m, and 2 m. The ALSE sounder

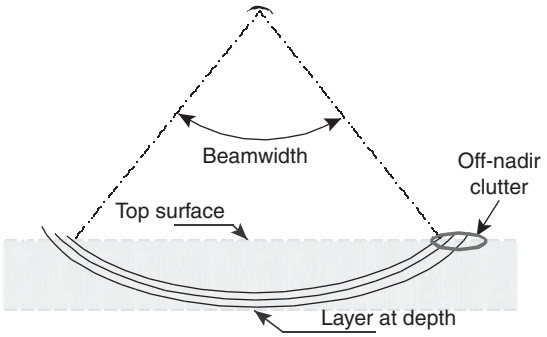


FIGURE 18.20 Two problems that challenge space-based subsurface sounding are (1) range-coincident competition between the desired return from depth and surface clutter and (2) range sidelobes from the very strong return from the top surface at nadir.

mode was designed to measure subsurface horizons in the lunar dielectric constant in both side-looking and down-looking viewing geometry. The instrument was based on synthetic aperture radar (SAR) principles. The data were recorded directly onto 70-mm photographic film, which was processed through a combination of optical and digital means following return of *Apollo 17* to Earth. (Optical processing was the state-of-the-art back in 1972 for SAR data.) Penetration depths proved to be approximately proportional to wavelength. The radar's waveform was constrained so that all sidelobes were at least 45 dB below the main-beam peak for all responses beyond three impulse widths of the compressed signal. Amplitude and phase linearity were constrained to be better than 0.1% and 0.001 radian, respectively. The transmitted pulses were linear FM, at 10% bandwidth in all three bands. Nominal free-space resolutions were 300 m, 100 m, and 10 m. Antenna gains were –0.8 dB, –0.7 dB, and +7.3 dB. These low numbers follow from the poor directivity of the antenna. Transmitted average power was 12 W, 4 W, and 1.5 W. Penetration depths were predicted to be 1300 m, 800 m, and 160 m, which turned out to be consistent with ALSE's performance.

MARSIS. The European Space Agency's *Mars Express* spacecraft includes the Mars Advanced Radar for Subsurface and Ionospheric Sounding instrument,[146] the first orbital sounder to fly since ALSE. MARSIS is a multifrequency down-looking radar that radiates 1-MHz pulses in one of four bands centered on the frequencies cited in Table 18.9. *Mars Express* is in an elliptical orbit; subsurface sounding operations are restricted to the lower 250 km to 800 km altitudes. Sounding is further limited by the ionosphere, which prevents effective radio wave propagation to the surface at frequencies below the plasma frequency f_0, which at Mars is ~4 MHz on the sunlit side and ~1 MHz on the dark side. The ionosphere induces a frequency-dependent time delay on the signal according to the index of refraction n, where $n = [1 - (f_0/f)^2]^{1/2}$. The resulting dispersion distorts the radar modulation, which must be compensated before pulse compression. The 40-m dipole antenna oriented orthogonal to the orbit plane is only weakly directive, with a 2.1 dB gain. The two 20-m long elements were not unfurled until two years into the mission, due to concerns about their potential for damaging the spacecraft during deployment. The effective cross-track footprint is on the order of 25 km. The along-track footprint is about 5 km, which is the result of onboard coherent doppler processing. This strategy reduces the impact of off-nadir clutter arising from sources in the along-track direction. The technique improves the subclutter visibility by 10 dB or more. The radar's mass and input power are 17 kg and 64 W. Peak radiated power is 10 W. The system noise floor is about 50 dB below the mean surface return, which establishes the dynamic range that limits depth of penetration. MARSIS has performed as intended,[147] with early results from the polar layered deposits, for example.

SHARAD. The Shallow Radar sounder[148] was designed to complement MARSIS. In general terms, it has higher resolution at a higher frequency, designed to provide sharper differentiation of the upper several hundred meters of the surface of Mars. SHARAD transmits a 10-MHz linear FM signal centered at 20 MHz. Theoretical vertical resolution is 7.5 m in material having a permittivity of 4. The 10-m dipole antenna has a frequency-dependent two-way gain of –5.7 dB to +0.2 dB. The effective beam footprint (after doppler processing) is ~0.5 km along track and ~5 km across track. Nominal SNR with respect to the surface return is in excess of 50 dB. Instrument mass is ~17 kg, input power is ~45 W. SHARAD measurements started late 2006 only when the MRO orbit was circularized after six months of aerobraking.

SELENE. The Japanese lunar mission, SELENE,[149] includes a 5 MHz Lunar Radar Sounder (LRS)[150] as one of its 14 payload instruments. The orbit is circular at 100-km altitude. The radar transmits 200 μsec linear FM signals that are demodulated using the stretch technique, similar to radar altimeters. Each pulse is amplitude-weighted by a sine function (0, π) to suppress sidelobes by ~30 dB from the surface return that otherwise would mask the desired returns from depth. The in situ resolution is nominally ~40 m within a medium of permittivity ~4. The radar dynamic range is ~50 dB to permit observation of subsurface profiles to a depth of several km. The antenna is comprised of two sets of dipoles, 30 m tip-to-tip, with an effective footprint of several tens of km. Peak output power is 800 W. Instrument mass is 24 kg and input power 50 W.

Atmospheric & Ionospheric Sounding–Flight Systems. The objective of this class of radar sounders is to generate a cross-section (profile) of the water or electron density in the orbit plane beneath the spacecraft. Atmospheric sounding by radar requires sensitivity to relatively weak backscatter, effective suppression (or avoidance) of the strong return from nadir, modest range resolution, and relatively narrow fields of view. These requirements lead to high power radars at K_u band or above, a simple pulse waveform, and substantial antenna area.

Radar sounders should not be confused with (passive) microwave radiometers—also called sounders, unfortunately—that are used by operational meteorological satellites to estimate atmospheric water distribution. A multifrequency radiometer generates coarse water vapor density profiles for which altitude is a function of frequency. Passive microwave sounding units have mass and power requirements on the order of 50 kg and 75 W, much less than their active radar counterparts.

TRMM. The Tropical Rainfall Measuring Mission was a joint undertaking between NASA and the Japanese Aerospace Exploration Agency (JAXA). The five-instrument payload includes the Precipitation Radar[151] (PR), which was designed and built by JAXA (then NASDA), and was the first of its kind on a space-based platform. Its precessing 400-km orbit, inclined at 35°, supports temporally and spatially sparse atmospheric coverage over tropical land and sea. The radar provides a three-dimensional structure of rainfall from the surface to an altitude of 20 km. When combined with data from the passive microwave radiometer (TMI), PR data support improved accuracy of rainfall retrievals. The radar's K_u band (~2 cm) frequency is about three times higher than most surface-based meteorological radars, but was selected to achieve a reasonably narrow 0.71° beamwidth from an antenna area constrained by spacecraft accommodation to be no more than 2.1 m by 2.1 m. Horizontal resolution at nadir is about 4.3 km. The antenna, a 128-element slotted waveguide phased array with 47 dB gain, is electronically scanned over ±17° across track, covering a swath 220-km wide, centered at nadir. Peak transmitted power is 500 W, generated from 128 solid-state power amplifiers (SSPAs), one for each waveguide. The 250-m range resolution is set by the unity time-bandwidth-product 1.6 μsec pulse. The radar has sufficient sensitivity to respond to rain rates as small as 0.7 mm/hr. Instrument mass is 465 kg; required input power is 250 W.

The ionospheric sounding mode on MARSIS (described above) is aimed primarily at characterization of the Martian ionosphere[152] during daylight conditions from altitudes below 1200 km. The radar is operated as a stepped-frequency instrument, sweeping 100 kHz to 5.4 MHz in 10.937 kHz intervals, over 7.38 seconds. From 500 km, the nominal SNR is 5.4 dB, increasing to 21.3 dB at 3.0 MHz.

CloudSat. Launched in April 2006, CloudSat includes as its primary payload the Cloud Profiling Radar[153] (CPR). CloudSat[154] flies in a sun-synchronous orbit in close formation with CALIPSO, which carries a cloud-profiling lidar, and in somewhat looser formation with Aqua, Aura, Terra, and PARASOL. Taken together, these six environmental satellites constitute the so-called A-Train. The average separation between CloudSat and CALIPSO is about 460 km, which corresponds to a one-minute delay between the radar and lidar cloud profiling measurements. The CPR was developed jointly by NASA and the Canadian Space Agency. It is a 94 GHz nadir-viewing real aperture radar, transmitting 3.3 μsec pulses at a PRF of 4.3 kHz to fill a window from the surface to 25 km altitude with 500-m vertical resolution sounding data. The antenna diameter, limited by the launch vehicle shroud, is 1.95 m, which supports across-track and along-track resolutions of 1.4 km and 2.5 km, respectively. The larger along-track resolution reflects the 0.3-sec integration time of the returns. Antenna gain is 63 dB. Average data rate is 15 kbits/s. Dynamic range is 70 dB; minimum detectable volumetric return is –26 dBZ. The PR mass is 230 kg; input power required is 270 W. Peak power transmitted is 1.7 kW.

The CPR's high-power amplifier (HPA) is the first of its kind for a space-based radar, an extended interaction klystron.[155] The klystron is driven by 200 mW signals from a solid-state preamplifier. The klystron requires 20 kV, provided by a high-voltage power supply system that also is a space first. The HPA is fully redundant.

DPR. The dual-frequency precipitation radar is the active microwave instrument[156] for the Global Precipitation Measurement (GPM) Core Observatory. DPR is based on a K_u band instrument (KuPR) similar to its predecessor on TRMM, augmented by a K_a band (35.55 GHz) radar (KaPR). Their two phased array slotted waveguide antennas are sized and oriented so that their footprints are the same. Their respective steered beams are synchronized so that for the central 100-km swath within which they both have coverage, their vertical profiles are near-simultaneous. The KuPR and KaPR antennas are sized at 2.4 m × 2.4 m and 1.4 m × 1.0 m, respectively, each comprised of 148 slotted waveguides driven by individual solid-state power amplifiers. Peak transmitted powers are 700 W and 140 W. Mass and input power of the two radars are 450 kg and 330 kg and 384 W and 326 W. The only other instrument in the payload is a microwave radiometer. To provide coverage at more latitudes, the spacecraft has a higher inclination orbit than TRMM's 39°. The main advantages of the second frequency and increased power are to distinguish between rain and snow and to increase sensitivity to rain rates as low as 0.2 mm/hr.

REFERENCES

1. D. E. Barrick and C. T. Swift, "The Seasat microwave instruments in historical perspective," *IEEE Journal of Oceanic Engineering*, vol. OE-5, pp. 74–79, 1980.
2. D. L. Evans, W. Alpers, A. Cazenave, C. Elachi, T. Farr, D. Glackin, B. Holt, L. Jones, W. T. Liu, W. McCandless, Y. Menard, R. Moore, and E. Njoku, "Seasat–A 25-year legacy of success," *Remote Sensing of Environment*, vol. 94, pp. 384–404, 2005.
3. M. D. Griffin and J. R. French, *Space Vehicle Design*, American Institute of Aeronautics and Astronautics, 2004.
4. V. J. Pisacane, *Fundamentals of Space Systems*, 2nd Ed, Oxford: Oxford University Press, 2005.
5. F. M. Henderson and A. J. Lewis (eds.), *Principles and Applications of Imaging Radar*, New York: J. Wiley & Sons Inc., 1998.

6. C. A. Wiley, "Synthetic Aperture Radars—A paradigm for technology evolution," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-21, pp. 440–443, 1985.
7. H. Jensen, L. C. Graham, L. J. Porcello, and E. N. Leith, "Side-looking airborne radar," *Scientific American*, vol. 237, pp. 84–95, 1977.
8. R. L. Jordan, "The Seasat-A synthetic-aperture radar system," *IEEE Journal of Oceanic Engineering*, vol. OE-5, pp. 154–164, 1980.
9. R. L. Jordan, B. L. Huneycutt, and M. Werner, "The SIR-C/X-SAR synthetic aperture radar system," *Proceedings of the IEEE*, vol. 79, pp. 827–838, 1991.
10. A. Freeman, M. Alves, B. Chapman, J. Cruz, Y. Kim, S. Shaffer, J. Sun, E. Turner, and K. Sarabandi, "SIR-C data quality and calibration results," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 848–857, 1995.
11. B. Rabus, M. Eineder, A. Roth, and R. Bamler, "The shuttle topography mission—a new class of digital elevation models acquired by spaceborne radar," *Photogrammetry and Remote Sensing*, vol. 57, pp. 241–262, 2003.
12. W. Yirong, Z. Minhui, and H. Wen, "SAR activities in P.R. China," in *Proceedings*, *6th European Conference on Synthetic Aperture Radar*, Dresden, Germany, VDE Verlag, 2006.
13. Y. Sharay and U. Naftaly, "TECSAR: design considerations and programme status," *IEE Proc. Radar Sonar Navigation*, vol. 153, pp. 117–121, 2006.
14. E. P. W. Attema, "The active microwave instrument on-board the ERS-1 satellite," *Proceedings of the IEEE*, vol. 79, pp. 791–799, 1991.
15. R. Bamler, M. Eineder, B. Kampes, H. Runge, and N. Adam, "SRTM and beyond: Current situation and new developments in spaceborne InSAR," in *Proceedings, ISPRS Workshop on High Resolution Mapping from Space*, Hanover, Germany, 2003.
16. Y. Nemoto, H. Nishino, M. Ono, H. Mizutamari, K. Nishikawa, and K. Tanaka, "Japanese earth resources satellite-1 synthetic aperture radar," *Proceedings of the IEEE*, vol. 79, pp. 800–809, 1991.
17. R. K. Raney, A. P. Luscombe, E. J. Langham, and S. Ahmed, "RADARSAT," *Proceedings of the IEEE*, vol. 79, pp. 839–849, 1991.
18. K. C. Jezek, K. Farness, R. Carande, X. Wu, and N. Labelle-Hamer, "RADARSAT-1 synthetic aperture radar observations of Antarctica: Modified Antarctic Mapping Mission, 2000," *Radio Science*, vol. 38, pp. 8067, 2003.
19. A. Ali, I. Barnard, P. A. Fox, P. Duggan, R. Gray, P. Allan, A. Brand, and R. Ste-Mari, "Description of RADARSAT-2 synthetic aperture radar design," *Canadian J. Remote Sensing*, vol. 30, pp. 246–257, 2004.
20. P. A. Fox, A. P. Luscombe, and A. A. Thompson, "RADARSAT-2 SAR modes development and utilization," *Canadian J. Remote Sensing*, vol. 30, pp. 258–264, 2004.
21. C. Zelli, "ENVISAT RA-2 Advanced radar altimeter: Instrument design and pre-launch performance assessment review," *Acta Astronautica*, vol. 44, pp. 323–333, 1999.
22. S. R. Cloude, G. Krieger, and K. P. Papathanassiou, "A framework for investigating space-borne polarimetric interferometry using the ALOS-PALSAR sensor," in *Proceedings IEEE Geoscience and Remote Sensing Symposium IGARSS2005*, Alaska, IEEE, 2005.
23. M. Shimada, M. Watanabe, T. Moriyama, and T. Tadono, "PALSAR characterization and initial calibration," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium IGARSS2006*, Denver, CO, IEEE, 2006.
24. A. Moreira, G. Krieger, D. Hounam, M. Werner, S. Riegger, and E. Settelmeyer, "TanDEM-X: A TerraSAR-X add-on satellite for single-pass SAR interferometry," in *Proc. International Geoscience and Remote Sensing Symposium IGARSS 2004*, Anchorage, Alaska, IEEE, 2004.
25. R. Levy-Nathansohn and U. Naftaly, "Overview of the TECSAR satellite modes of operation," in *Proceedings*, *6th European Conference on Synthetic Aperture Radar*, Dresden, Germany, VDE Verlag, 2006.
26. Y. Wu, M. Zhu, and W. Hong, "SAR activities in P. R. China," in *Proceedings of EUSAR*, Dresden, Germany, ITG VDE, 2006.

27. T. Misra, S. S. Rana, V. H. Bora, N. M. Desai, C. V. N. Rao, and N. Rajeevjyothi, "SAR payload of radar imaging satellite RISAT) of ISRO," in *Proceedings*, *6th European Conference on Synthetic Aperture Radar*, Dresden, Germany, VDE Verlag, 2006.
28. R. Schroeder, J. Puls, F. Jochim, J.-L. Bueso-Bello, L. Datashvili, H. Baier, M. M. Quintino da Silve, and W. Paradella, "The MAPSAR mission: Objectives, design, and status," in *Proceedings of EUSAR*, Dresden, Germany, ITG VDE, 2006.
29. H. Kimura and N. Itoh, "ALOS PALSAR: The Japanese second-generation spaceborne SAR and its application," *Proc. Society of Photo-optical Instrumentation Engineers (SPIE)*, vol. 4152, pp. 110–119, 2000.
30. J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, New York: John Wiley & Sons, Inc., 1991.
31. R. K. Raney, "Radar fundamentals: technical perspective," in *Principles and Applications of Imaging Radar*, F. Henderson and A. Lewis (eds.), New York: Wiley Interscience, 1998, pp. 9–130.
32. G. Franceschetti and R. Lanari, *Synthetic Aperture Radar Processing*, Boca Raton, FL: CRC Press, 1999.
33. R. K. Raney, "Considerations for SAR image quantification unique to orbital systems," *IEEE Transactions Geoscience and Remote Sensing*, vol. 29, pp. 754–760, 1991.
34. P. E. Green, Jr., "Radar measurements of target scattering properties," In *Radar Astronomy*, J. V. Evans and T. Hagfors (eds.), New York: McGraw-Hill, 1968.
35. W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar—Signal Processing Algorithms*, Boston: Artech House, 1995.
36. C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Boston: Kluwer Academic Publishers, 1996.
37. R. Bamler, "Optimum look weighting for burst-mode and scanSAR processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 722–725, 1995.
38. K. Tomiyasu, "Conceptual performance of a satellite-borne, wide swath synthetic aperture radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 19, pp. 108–116, 1981.
39. R. K. Moore, J. P. Claasen, and Y. H. Lin, "Scanning spaceborne synthetic aperture radar with integrated radiometer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-17, pp. 410–421, 1981.
40. A. Luscombe, A. Thompson, P. James, and P. Fox, "Calibration techniques for the RADARSAT-2 SAR system," in *Proceedings of EUSAR 2006*, Dresden, Germany, VDE Verlag, 2006.
41. W. M. Boerner, H. Mott, E. Luneburg, C. Livingstone, B. Brisco, R. J. Brown, and J. S. Paterson, "Polarimetry in radar remote sensing: basic and applied concepts," in *Principles and Applications of Imaging Radar*, F. M. Henderson and A. J. Lewis (eds.), New York: John Wiley & Sons, Inc., 1998.
42. L. Graham, "Synthetic interferometers for topographic mapping," *Proc. IEEE*, vol. 62, pp. 763–768, 1974.
43. R. Gens and J. vanGenderen, "Review article: SAR interferometry—issues, techniques, applications," *Int. J. Remote Sensing*, vol. 17, pp. 1803–1835, 1996.
44. P. Rosen, S. Hensley, I. Joughin, F. Li, S. Madsen, E. Rodriguez, and R. Goldstein, "Synthetic aperture radar interferometry," *Proc. IEEE*, vol. 88, pp. 333–382, 2000.
45. R. F. Hanssen, *Radar Interferometry*, Dordrecht, The Netherlands: Klewer Academic Publishers, 2001.
46. H. Zebker and R. Goldstein, "Topographic mapping from interferometric synthetic aperture radar observations," *J. Geophys. Res.*, vol. 91, pp. 4993–4999, 1986.
47. S. Madsen, H. Zebker, and J. Martin, "Topographic mapping using radar interferometry," *IEEE Trans Geoscience and Remote Sensing*, vol. 31, pp. 246–256, 1993.
48. A. Gabriel, R. Goldstein, and H. Zebker, "Mapping small elevation changes over large areas: Differential radar interferometry," *J. Geophys. Res.*, vol. 94, pp. 9183–9191, 1989.
49. D. Massonnet and K. Feigl, "Radar interferometry and its application to changes in the Earth's surface," *Rev. Geophysics*, vol. 36, pp. 441–500, 1998.

50. M. Born and E. Wolf, *Principles of Optics*, New York: Pergamon Press, Macmillan, 1959.
51. F. Gatelli, A. Guarnieri, F. Parizzi, P. Pasquali, C. Prati, and F. Rocca, "The wavenumber shift in SAR interferometry," *IEEE Trans Geoscience and Remote Sensing*, vol. 32, pp. 855–865, 1994.
52. H. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans Geoscience and Remote Sensing*, vol. 30, pp. 950–959, 1992.
53. A. Ferretti, C. Prati, and F. Rocca, "Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry," *IEEE Trans Geoscience and Remote Sensing*, vol. 38, pp. 2202–2212, 2000.
54. T. H. Dixon, F. Amelung, A. Ferretti, F. Novali, F. Rocca, R. Dokka, G. Sella, S.-W. Kim, S. Wdowinski, and D. Whitman, "Space geodesy: Subsidence and flooding in New Orleans," *Nature*, vol. 441, pp. 587–588, 2006.
55. D. Ghiglia and M. Pritt, *Two-dimensional Phase Unwrapping: Theory, Algorithms, and Software*, New York: Wiley, 1998.
56. R. Goldstein and C. Werner, "Radar interferogram filtering for geophysical applications," *Geophysical Res. Letters*, vol. 25, pp. 4035—4038, 1998.
57. J. J. van Zyl, H. A. Zebker, and C. Elachi, "Imaging radar polarization signatures: Theory and observation," *Radio Science*, vol. 22, pp. 529–543, 1987.
58. P. E. Green Jr., "Radar measurements of target scattering properties," in *Radar Astronomy*, J. V. Evans and T. Hagfors (eds.), New York: McGraw-Hill Book Company, 1968, pp. 1–78.
59. A. Guissard, "Meuller and Kennaugh matrices in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 590–597, 1994.
60. C. Lopez-Martinez, E. Pottier, and S. R. Cloude, "Statistical assessment of Eigenvector-based target decomposition theorems in radar polarimetry," *IEEE Trans. Geoscience and Remote Sensing*, vol. 43, pp. 2058–2074, 2005.
61. S. R. Cloude and K. P. Papathanassiou, "Polarimetric SAR interferometry," *IEEE Trans Geoscience and Remote Sensing*, vol. 36, pp. 1551–1565, 1998.
62. S. R. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geoscience and Remote Sensing*, vol. 35, pp. 68–78, 1997.
63. L.-L. Fu and A. Cazanave (eds.), *Satellite Altimetry and the Earth Sciences*, San Diego: Academic Press, 2001.
64. J. R. Jensen and R. K. Raney, "Delay Doppler radar altimeter: Better measurement precision," in *Proceedings IEEE Geoscience and Remote Sensing Symposium IGARSS'98*, Seattle, WA, IEEE, 1998, pp. 2011–2013.
65. R. K. Moore and C. S. Williams, Jr., "Radar return at near-vertical incidence," *Proceedings of the IRE*, vol. 45, pp. 228–238, 1957.
66. G. S. Brown, "The average impulse response of a rough surface and its applications, *IEEE Antennas and Propagation*, vol. 25, pp. 67–74, 1977.
67. E. J. Walsh, "Pulse-to-pulse correlation in satellite radar altimetry," *Radio Science*, vol. 17, pp. 786–800, 1982.
68. J. T. McGoogan, L. S. Miller, G. S. Brown, and G. S. Hayne, "The S-193 radar altimeter experiment," *Proceedings of the IEEE*, vol. 62, pp. 793–803, 1974.
69. G. S. Hayne, "Radar altimeter mean return waveforms from near-normal incidence ocean surface scattering," *IEEE Antennas and Propagation*, vol. AP-28, pp. 687–692, 1980.
70. J. L. MacArthur, C. C. Kilgus, C. A. Twigg, and P. V. K. Brown, "Evolution of the satellite radar altimeter," *Johns Hopkins APL Technical Digest*, vol. 10, pp. 405–413, October–December 1989.
71. J. L. MacArthur, P. C. Marth, and J. G. Wall, "The GEOSAT radar altimeter," *Johns Hopkins APL Technical Digest*, vol. 8, pp. 176–181, 1987.
72. W. H. F. Smith and D. T. Sandwell, "Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry," *J. Geophys. Res.*, vol. 99, pp. 21803–21824, 1994.
73. D. T. Sandwell and W. H. F. Smith, "Bathymetric estimation," in *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave (eds.), New York: Academic Press, 2001, pp. 441–457.
74. APL, Special sections, "Geosat science and altimeter technology," *Johns Hopkins APL Technical Digest*, vol. 10, 1989.

75. D. B. Chelton, J. C. Ries, B. J. Haines, L.-L. Fu, and P. S. Callahan, "Satellite altimetry," in *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazanave (eds.), San Diego: Academic Press, 2001, pp. 1–122.
76. J. Goldhirsh and J. R. Rowland, "A tutorial assessment of atmospheric height uncertainties for high-precision satellite altimeter missions to monitor ocean currents," *IEEE Transactions Geoscience and Remote Sensing*, vol. 20, pp. 418–434, 1982.
77. D. J. Wingham, l. Phalippou, C. Mavrocordatos, and D. Wallis, "The mean echo and echo cross-product from a beamforming interferometric altimeter and their application to elevation measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, pp. 2305–2323, 2004.
78. P. Vincent, N. Steunou, E. Caubet, L. Phalippou, L. Rey, E. Thouvenot, and J. Verron, "AltiKa: a Ka-band altimetery payload and system for operational altimetry during the GMES period," *Sensors*, vol. 6, pp. 208–234, 2006.
79. M. E. Parke, R. H. Stewart, D. L. Farless, and D. E. Cartwright, "On the choice of orbits for an altimetric satellite to study ocean circulation and tides," *Journal of Geophysical Research*, vol. 92, pp. 11693–11707, October 15, 1987.
80. D. T. Sandwell and W. H. F. Smith, "Marine gravity anomaly from Geosat and ERS-1 satellite altimetry," *J. Geophys. Res.*, vol. 102, pp. 10039–10054, 1997.
81. R. D. Ray, "Applications of high-resolution ocean topography to ocean tides," in *Report of the High-Resolution Ocean Topography Science Working Group Meeting*, D. Chelton (ed.), Corvallis, Oregon: Oregon State University, 2001.
82. R. Scharroo and P. Visser, "Precise orbit determination and gravity field improvement for the ERS satellites," *J. of Geophysical Research*, vol. 103, pp. 8113–8127, 1998.
83. A. R. Zieger, D. W. Hancock, G. S. Hayne, and C. L. Purdy, "NASA radar altimeter for the TOPEX/Poseidon project," *Proceedings of the IEEE*, vol. 79, pp. 810–826, June 1991.
84. P. C. Marth, J. R. Jensen, C. C. Kilgus, J. A. Perschy, J. L. MacArthur, D. W. Hancock, G. S. Hayne, C. L. Purdy, L. C. Rossi, and C. J. Koblinsky, "Prelaunch performance of the NASA altimeter for the TOPEX/Poseidon Project," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, pp. 315–332, 1993.
85. D. B. Chelton, E. J. Walsh, and J. L. MacArthur, "Pulse compression and sea-level tracking in satellite altimetry," *Journal of Atmospheric and Oceanic Technology*, vol. 6, pp. 407–438, 1989.
86. W. J. J. Caputi, "Stretch: a time-transformation technique," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-7, pp. 269–278, 1971.
87. D. T. Sandwell, "Antarctic marine gravity field from high-density satellite altimetry," *Geophys. J. Int.*, vol. 109, pp. 437–448, 1992.
88. W. H. F. Smith and D. T. Sandwell, "Conventional bathymetry, bathymetry from space, and geodetic altimetry," *Oceanography*, vol. 17, pp. 8–23, 2004.
89. M. M. Yale, D. T. Sandwell, and W. H. F. Smith, "Comparison of along-track resolution of stacked Geosat, ERS-1 and TOPEX satellite altimeters," *J. Geophys. Res.*, vol. 100, pp. 15117–15127, 1995.
90. L. Phalippou, L. Rey, P. DeChateau-Thierry, E. Thouvenot, N. Steunou, C. Mavrocordatos, and R. Francis, "Overview of the performances and tracking design of the SIRAL altimeter for the CryoSat mission," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, pp. 2025–2027, 2001.
91. R. K. Raney, "The delay doppler radar altimeter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp. 1578–1588, 1998.
92. J. R. Jensen, "Design and performance analysis of a phase-monopulse radar altimeter for continental ice sheet monitoring," in *Proceedings, IEEE International Geoscience and Remote Sensing Symposium IGARSS'95*, Florence, Italy, IEEE, 1995, pp. 865–867.
93. R. K. Raney and J. R. Jensen, "An Airborne CryoSat Prototype: The D2P Radar Altimeter," in *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'02*, Toronto, IEEE, 2002, pp. 1765–1767.
94. S. Laxon, N. Peacock, and D. Smith, "High interannual variability of sea ice thickness in the Arctic region," *Letters to Nature*, vol. 425, pp. 947–950, 2003.

95. A. J. Butrica, *To See the Unseen: A History of Planetary Radar*, Darby, PA: Diane Publications, 1997.
96. S. J. Ostro, "Planetary radar astronomy," in *The Encyclopedia of Physical Science and Technology, 3rd Edition*, R. A. Meyers (ed.), San Diego, Academic Press, 2002, pp. 295–328.
97. D. B. Campbell, R. S. Hudson, and J.-L. Margot, "Advances in planetary radar astronomy," Chapter 35, in *Review of Radio Science, 1999–2002*, R. Stone (ed.), Oxford: U.R.S.I, 2002, pp. 869–899.
98. S. Slavney, R. E. Arvidson, K. Bennett, E. A. Guiness, and T. C. Stein, "Recent and planned Planetary Data System geosciences node activities," Paper 2232.pdf, *Proceedings, Lunar and Planetary Science XXXVII*, Houston, TX, vol., 2006.
99. G. H. Pettengill, P. G. Ford, and B. D. Chapman, "Venus: surface electromagnetic properties," *J. Geophys. Res.*, vol. 93, pp. 14881-14892, 1988.
100. B. A. Ivanov, "Venusian impact craters on Magellan images: View from Venera 15/16," *Earth Moon Planet*, vol. 50/51, pp. 159–173, 1990.
101. G. H. Pettengill, P. G. Ford, W. T. K. Johnson, R. K. Raney, and L. A. Soderblom, "Magellan: Radar performance and data products," *Science*, vol. 252, pp. 260–265, 1991.
102. W. T. K. Johnson, "Magellan imaging radar mission to Venus," *Proceedings of the IEEE*, vol. 79, pp. 777–790, 1991.
103. C. Elachi, E. Im, L. E. Roth, and C. L. Werner, "Cassini Titan radar mapper," *Proceedings of the IEEE*, vol. 79, pp. 867–880, 1991.
104. S. Nozette, C. L. Lichtenberg, P. Spudis, R. Bonner, W. Ort, E. Malaret, M. Robinson, and E. M. Shoemaker, "The Clementine bistatic radar experiment," *Science*, vol. 274, pp. 1495–1498, November 29, 1996.
105. B. Hapke, "Coherent backscatter and the radar characteristics of outer planet satellites," *Icarus*, vol. 88, pp. 407–417, December 1990.
106. R. A. Simpson and G. L. Tyler, "Reanalysis of Clementine bistatic radar data from the lunar South Pole," *J. Geophys. Res.*, vol. 104, pp. 3845–3862, February 25, 1999.
107. D. B. Campbell, B. A. Campbell, L. M. Carter, J.-L. Margot, and N. J. S. Stacy, "No evidence for thick deposits of ice at the lunar south pole," *Nature*, vol. 443, pp. 835–837, 2006.
108. M. T. Zuber and I. Garrick-Bethell, "What do we need to know to land on the Moon again?," *Science*, vol. 310, pp. 983–985, 2005.
109. K. J. Peters, "Coherent-backscatter effect: A vector formulation accounting for polarization and absorption effects and small or large scatterers," *Physical Review B*, vol. 46, pp. 801–812, 1 July 1992.
110. P. D. Spudis, C. L. Lichtenberg, B. Marinelli, and S. Nozette, "Mini-SAR: An imaging radar for the Chandrayaan-1 mission to the Moon," Paper 1153, *Proceedings, Lunar and Planetary Science XXXVI*, Houston, TX, vol., 2005.
111. R. K. Raney, "Hybrid-polarity SAR architecture," in *Proceedings IEEE Geoscience and Remote Sensing Symposium*, Denver, CO, IEEE, 2006.
112. D. J. Mudgway, *Big Dish: Building America's Deep Space Connection to the Planets*, Gainesville: University of Florida Press, 2005.
113. R. K. Moore, "Trade-off between picture element dimensions and noncoherent averaging in side-looking airborne radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 15, pp. 697–708, 1979.
114. R. K. Raney, "The making of a precedent: the synthetic aperture radar (SAR) on Magellan," *V-GRAM (Jet Propulsion Laboratory)*, vol. 9, pp. 3–10, 1986.
115. R. Kwok and W. T. K. Johnson, "Block adaptive quantization of Magellan SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 27, pp. 375–383, 1989.
116. G. G. Stokes, "On the composition and resolution of streams of polarized light from different sources," *Transactions of the Cambridge Philosophical Society*, vol. 9, pp. 399–416, 1852.
117. S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geoscience and Remote Sensing*, vol. 34, pp. 498–518, 1996.
118. R. K. Raney, "Stokes parameters and hybrid-polarity SAR architecture," *IEEE Transactions Geoscience and Remote Sensing*, vol. 45, pp. 3397–3404, 2007.

119. Y. Quilfen, B. Chapron, F. Collard, and D. Vandemark, "Relationship between ERS scatterometer measurement and integrated wind and wave parameters," *J. Atmospheric and Oceanic Technology*, vol. 21, pp. 368–373, 2004.
120. R. K. Moore and A. K. Fung, "Radar determination of winds at sea," *Proceedings of the IEEE*, vol. 67, pp. 1504–1521, 1979.
121. J. Kerkmann, *Review on Scatterometer Winds*, Darmstadt, Germany: European Organization for the Exploitation of Meteorological Satellites EUMETSAT, 1998, p. 77.
122. A. Mouche, D. Hauser, and V. Kudryavstev, "Observations and modelling of the ocean radar backscatter at C-band in HH- and VV-polarizations," in *Proceedings International Geoscience and Remote Sensing Symposium*, Seoul, Korea, IEEE, 2005.
123. M. Migliaccio, "Sea wind field retrieval by means of microwave sensors: a review," in *Proc. URSI Commission F Symposium*, Ispra, Italy, 2005.
124. E. M. Bracalente, D. H. Boggs, W. L. Grantham, and J. L. Sweet, "The SASS scattering coefficient ó0 algorithm," *IEEE Journal of Oceanic Engineering*, vol. OE-5, pp. 145–154, 1980.
125. R. E. Fischer, "Standard deviation of scatterometer measurements from space," *IEEE Transactions on Geoscience Electronics*, vol. GE-10, pp. 106–113, 1972.
126. P. W. Gaiser and C. S. Ruf, "Foreword to the special issue on the WindSat Spaceborne Polarimetric Radiometer—calibration/validation and wind vector retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, pp. 467–469, 2006.
127. I. H. Woodhouse and D. H. Hoekman, "Determining land-surface parameters from the ERS wind scatterometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 126–140, 2000.
128. D. G. Long, M. R. Drinkwater, B. Holt, S. Saatchi, and C. Bertoia, "Global ice and land climate studies using scatterometer image data," *EOS, Trans. American Geophysical Union*, vol. 82, pp. 503, 2001.
129. M. R. Drinkwater, D. G. Long, and A. W. Bingham, "Greenland snow accumulation estimates from satellite radar scatterometer data," *J. of Geophysical Research*, vol. 106(D24), pp. 33935–33950, 2001.
130. L. B. Kunz and D. G. Long, "Calibrating SeaWinds and QuikSCAT scatterometers using natural land targets," *IEEE Geoscience and Remote Sensing Letters*, vol. 2, pp. 182–186, 2005.
131. R. K. Moore, "Simultaneous active and passive microwave response of the Earth: The Skylab RADSCAT experiment," in *Proceedings of the 9th International Symposium on Remote Sensing*, Ann Arbor, Michigan, 1974, pp. 189–217.
132. M. Shimada and A. Freeman, "A technique for measurement of spaceborne SAR antenna patterns using distributed targets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 100–114, 1995.
133. W. L. Grantham, E. M. Bracalente, L. W. Jones, and J. W. Johnson, "The Seasat-A satellite scatterometer," *IEEE Journal of Oceanic Engineering*, vol. OE-2, pp. 200–206, 1977.
134. J. W. Johnson, L. A. Williams, Jr., E. M. Bracalente, F. B. Beck, and W. L. Grantham, "Seasat-A satellite scatterometer instrument evaluation," *IEEE Journal of Oceanic Engineering*, vol. OE-5, pp. 138–144, 1980.
135. J. Figa-Saldana, J. J. W. Wilson, E. Attema, R. Gelsthorpe, M. R. Drinkwater, and A. Stoffelen, "The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: a follow-on for European wind scatterometers," *Canadian J. of Remote Sensing*, vol. 28, pp. 404–412, 2002.
136. F. M. Naderi, M. H. Freilich, and D. G. Long, "Spaceborne radar measurement of wind velocity over the ocean—An overview of the NSCAT scatterometer system," *Proceedings of the IEEE*, vol. 79, pp. 850–866, 1991.
137. C. Wu, J. Graf, M. H. Freilich, D. G. Long, M. W. Spencer, W.-Y. Tsai, D. Lisman, and C. Winn, "The SeaWinds scatterometer instrument," in *Proceedings, IEEE International Geoscience and Remote Sensing Symposium*, Pasadena, CA, pp. 1511–1515, 1994.
138. M. W. Spencer, C. Wu, and D. G. Long, "Tradeoffs in the design of a spaceborne scanning pencil beam scatterometer: Application to Sea-Winds," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 115–126, 1997.

139. C. Wu, Y. Liu, K. H. Kellogg, K. S. Pak, and R. L. Glenister, "Design and calibration of the SeaWinds scatterometer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, pp. 94–109, 2003.

140. M. W. Spencer, C. Wu, and D. G. Long, "Improved resolution backscatter measurements with the SeaWinds pencil-beam scatterometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 89–104, 2000.

141. S. H. Yueh, "Microwave remote sensing modeling of ocean surface salinity and winds using an empirical sea surface spectrum," in *Proceedings IEEE Geoscience and Remote Sensing Symposium*, Anchorage, AK, 2004.

142. S. Gogineni, D. Tammana, D. Braaten, C. Leuschen, T. Atkins, J. Legarsky, P. Kanagaratnam, J. Stiles, C. Allen, and K. Jezek, "Coherent radar ice thickness measurements over Greenland ice sheet," *Journal of Geophysical Research*, vol. 106, pp. 33761–33772, 2001.

143. S. H. Ward, G. R. Jiracek, and W. I. Linlor, "Electromagnetic reflection from a plane-layered lunar model," *Journal of Geophysical Research*, vol. 73, pp. 1355–1372, 1968.

144. L. J. Porcello, R. L. Jordan, J. S. Zelenka, G. F. Adams, R. J. Phillips, W. E. Brown, Jr., S. H. Ward, and P. L. Jackson, "The Apollo lunar sounder radar system," *Proceedings of the IEEE*, vol. 62, pp. 769–783, 1974.

145. W. J. Peeples, W. R. Sill, T. W. May, S. H. Ward, R. J. Phillips, R. L. Jordan, E. A. Abbott, and T. J. Killpack, "Orbital radar evidence for lunar subsurface layering in Maria Serenitatis and Crisium," *J. of Geophysical Research*, vol. 83, pp. 3459–3468, 1978.

146. D. Biccari, F. Ciabattoni, G. Picardi, R. Seu, W. T. K. Johnson, R. L. Jordan, J. Plaut, A. Safaeinili, D. A. Gurnett, R. Orosei, O. Bombaci, F. Provvedi, and E. Zampolini, "Mars advanced radar for subsurface and ionosphere sounding (MARSIS)," in *Proc. 2001 International Conference on Radar*, Beijing, China, 2001.

147. J. Farrell, J. Plaut, A. Gurnett, and G. Picardi, "Detecting sub-glacial aquifers in the North Polar layered deposits with Mars Express/MARSIS," *Geophysical Research Letters*, vol. 32, pp. L11204, June 10, 2005.

148. M. Kato, Y. Takizawa, S. Sasaki, and the SELENE Project Team, "SELENE, the Japanese lunar orbiting satellite mission: present status and science goals," in *Proceedings, Lunar and Planetary Science XXXVII*, Houston, TX, pp. 1233.pdf, 2006.

149. T. Ono, T. Kobayashi, and H. Oya, "Interim report of the Lunar Radar Sounder on-board SELENE spacecraft," in *Proceedings, 35th COSPAR Assembly*, Paris, France, pp. 3315, 2004.

150. E. Im, E., S. L. Durden, S. Tanelli, and K. Pak, "Early results on cloud profiling radar post-launch calibration and operations," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Denver, CO, 2006.

151. G. L. Stephens and D. G. Vane, "The CloudSat mission," in *Proceedings IEEE International Geoscience and Remote Sensing Symposium*, Toulouse, France, 2003.

152. A. Roitman, D. Berry, and B. Steer, "State-of-the-art W-band extended interaction klystron for the CloudSat program," *IEEE Transactions on Electron Devices*, vol. 52, pp. 895–898, 2005.

153. Y. Senbokuva, S. Satoh, K. Furukawa, M. Koiima, H. Hanado, N. Takahashi, T. Iquchi, and K. Nakamura, "Development of the spaceborne dual-frequency precipitation radar for the Global Precipitation Measurement mission," in *Proceedings International Geoscience and Remote Sensing Symposium*, Anchorage, Alaska, pp. 3566–3569, 2004.

154. R. Seu, D. Biccari, R. Orosei, L. V. Lorenzoni, R. J. Phillips, L. Marinangeli, G. Picardi, A. Masdea, and E. Zampolini, "SHARAD: the MRO 2005 shallow radar," *Planetary and Space Science*, vol. 52, pp. 157–166, 2004.

155. T. Kozu, T. Kawanishi, H. Kuroiwa, M. Kojima, K. Oikawa, H. Kumagai, K. Okamoto, M. Okamura, H. Nakatuka, and K. Nishikawa, "Development of precipitation radar onboard the Tropical Rainfall Measurement Mission (TRMM) satellite," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 102–116, 2001.

156. D. A. Gurnett, D. L. Kirchner, R. L. Huff, D. D. Morgan, A. M. Persoon, T. F. Averkamp, F. Duru, E. Nielsen, A. Safaeinili, J. J. Plaut, and G. Picardi, "Radar soundings of the ionosphere of Mars," *Science*, vol. 310, pp. 1929–1933, December 23, 2005.

Chapter 19
Meteorological Radar

R. Jeffrey Keeler and Robert J. Serafin

*National Center for Atmospheric Research**

19.1 INTRODUCTION

Standard operational meteorological doppler radars have become familiar observational tools to radar engineers as well as the general public since their introduction by the United States National Weather Service (NWS) in the 1950s and are widely used by weather forecasters in the public and private sectors. Major technical improvements were introduced in the 1990s when the NWS, the Federal Aviation Agency (FAA), and the U.S. Air Force joined together to install the next generation national network of WSR-88D doppler radars (commonly called Nexrad radars). Also in the 1990s the FAA installed the Terminal Doppler Weather Radar (TDWR) system at major airports in the U.S. The federal agencies have subsequently implemented many technical upgrades to the Nexrad and TDWR networks to improve their performance for public warnings and aviation safety.[1,2] In contrast to the WSR-57 and WSR-74C radars they replaced, the WSR-88D systems provide quantitative and automated real-time information on storms, precipitation, hurricanes, tornadoes, and a host of other important weather phenomena with higher spatial and temporal resolution than ever before.[3,4] In the aviation community, TDWR radars provide crucial information for providing safe departures and landings at the major airports by detecting hazardous wind shear events such as microbursts, strong gust fronts, and other performance reducing wind hazards.[5,6,7]

Various other types of meteorological radars exist. The Nexrad long-range weather surveillance radars are frequently supplemented by (typically) smaller medium-range weather radars operated by TV stations for local observations.[8] In addition to the familiar commercial airborne weather avoidance and observation radars, airborne hurricane monitoring provides detailed forecasts and warnings for approaching coastal hurricanes.[9] And vertical pointing wind profiling fixed-beam systems are routinely used to obtain continuous profiles of horizontal winds[10] whereas spaced-based meteorological radars are measuring widespread equatorial precipitation fields and cloud properties.[11] Meteorological research results are regularly transferred to the operational weather radar community for achieving higher space and time resolution, for improved data quality, and for the production of new weather radar products all of which have led to dramatic improvements in weather forecasting. Doppler weather radars measure detailed vector wind fields as well as precipitation fields. Small, highly mobile research radars provide many of the same capabilities as the fixed radars.[12] Dual polarization techniques[13,14] are used for improved quantitative

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

precipitation measurement, for detecting hail,[15] and for discriminating ice particles (snow) from water (rain).[16] Furthermore, ground-based research radars can now measure atmospheric moisture in the surface boundary layer.[17] Airborne research radars provide many of these same capabilities with increased coverage and greater mobility.[18] This variety of applications in both research and operations illustrates the vitality of meteorological radar technology and its evolution.

This chapter is intended to introduce the reader to meteorological radar, particularly those system characteristics that are unique to meteorological applications. In this regard, it should be noted that most meteorological radars appear similar to radars used for other purposes. Pulsed doppler systems are far more prevalent than CW radars. Primarily, center-fed parabolic dish antennas with focal-point feed and low-noise solid-state digital receivers are used. Magnetrons, klystrons, traveling-wave tubes, and other forms of transmitter forms are routinely used.

The distinguishing factor between meteorological radar and other kinds of aviation or military radars lies in the nature of weather targets, the resulting characteristics of the radar signal, and the means by which these weather echoes are processed to suppress artifacts and generate only the significant and essential weather information. Important meteorological targets occupy a wide range of scattering echo intensities (−20 to 70 dBZ) that are distributed in space from short range (< 1 km) to long range (> 200 km), close to the surface (100 m) to the top of the atmosphere where weather is important (20 km), and typically occupy a large fraction of the several millions of spatial resolution cells observed by the radar. Moreover, it is necessary to make quantitative measurements of the received signal characteristics in each of these cells, or "weather targets," to estimate such parameters as precipitation rate, precipitation type, air motion, turbulence, and wind shear.[19] In addition, because a high percentage of radar resolution cells contain useful information, meteorological radars require fast digital signal processors, effective means for suppressing artifacts caused by the data density, high data-rate recording systems, and informative displays of this information. Thus, whereas many non-weather radar applications call for detection, tracking, and detailed characterization of relatively fewer number of desired targets in a field of widespread weather, ground, sea, decoy, and bird clutter, the meteorological radars focus on making accurate estimates of the nature of the "weather clutter" itself. Both aviation/military and weather radars require heavy processing activity, but the volume of data for assimilation, recording, and display of weather radars is often much larger since different essential information must be extracted for a large number of anticipated users when measuring widespread weather systems.

The discussion herein refers to a number of useful texts and references for the reader. However, the classic *Radar Observation of the Atmosphere* by Lou Battan[20] deserves special mention for its clarity and completeness and remains a standard for courses in radar meteorology. The *Battan Memorial and 40th Anniversary Radar Meteorology Conference* produced a collection of review papers, *Radar in Meteorology,*[21] covering the first four decades of radar meteorology from the historical, technological, scientific, and operational perspectives. Bean et al.[22] in Skolnik's first *Radar Handbook* addressed the problem of weather effects on radar. Doviak and Zrnić[23] place special emphasis on doppler aspects of meteorological radar, whereas Bringi and Chandra[24] emphasize all aspects of polarimetric radars and Lhermitte[25] focuses on millimeter wave (cloud) radars. Rinehart's *Radar for Meteorologists*[26] gives a broad and easily comprehendible overview of all aspects of weather radar. The IEEE Geoscience and Electronics *Special Issue on Radar Meteorology,*[27] Atlas's *Radar in Meteorology,*[21] Wakimoto and Srivastiva's *Radar and Atmospheric Science: A Collection of Essays in Honor of David Atlas,*[28] and

Meischner's *Weather Radar*[29] provide an evolving perspective on many aspects of meteorological radars by technological and scientific leaders. Finally, perhaps the broadest and most complete set of references on progress in the field can be found in the series of *Proceedings and Preprints of the (International) Conferences on Radar Meteorology*[30] sponsored by the American Meteorological Society (AMS). These documents can be found in many technical libraries and also can be obtained online. In addition, the *Proceedings of the European Conferences on Radar Meteorology*[31] provides excellent reference material.

19.2 THE RADAR EQUATION FOR METEOROLOGICAL TARGETS

The received power P_r from a radar point target can be derived from any of a variety of expressions that are applicable to radar in general.[23,26,32] For a single point target, a simple form that is readily derived is

$$P_r = \frac{\beta\sigma}{r^4} \tag{19.1}$$

where β is a constant dependent upon radar system parameters (transmitted power P_t, antenna system gain G, and wavelength λ), r is the range to the point target, and σ is the radar cross section (RCS).*

It is in the calculation of σ for distributed meteorological targets that the radar equation differs from that for point targets. For distributed targets like rainfall the RCS may be written

$$\sigma = \eta V \tag{19.2}$$

where η is the radar reflectivity in units of cross-sectional area per unit volume and V is the volume sampled by the radar. η can itself be written as

$$\eta = \sum_{i=1}^{N} \sigma_i \tag{19.3}$$

where N is the number of scatterers per unit volume and σ_i is the backscattering cross section of the ith point scatterer. In general, the meteorological scatterers can take on a variety of forms, which include water droplets, ice crystals, hail, snow, and mixtures of the above.

Mie[33] developed a general theory for the energy backscattered by an optical plane wave impinging on conducting spheres in colloidal suspension. The same theory applies to spherical raindrops falling through the atmosphere for which the backscattered energy is a function of the wavelength (λ) of the incident energy and the radius (a) and complex index of refraction (m) of the particle. The ratio $2\pi a/\lambda$ determines the dominant scattering properties of the particle. Spherical water droplets in air that are large relative to the wavelength scatter in the so called optical region; droplets on the order of the same size as the wavelength scatter in the so called resonant scattering region; and droplets small relative to the wavelength scatter in the so-called Rayleigh region.

* By convention in this chapter, we shall use r for range and R for rainfall rate.

When the ratio $2\pi a/\lambda < 1$, the Rayleigh approximation[20] may be applied, and σ_i becomes

$$\sigma_i = \frac{\pi^5}{\lambda^4} |K|^2 D_i^6 \tag{19.4}$$

where D_i is the diameter of the ith drop and

$$|K|^2 = \left|\frac{m^2 - 1}{m^2 + 2}\right|^2 \tag{19.5}$$

where m is the complex index of refraction. At temperatures between 0 and 20°C and centimeter wavelengths, $|K|^2 \approx 0.93$ for the water phase and $|K|^2 \approx 0.197$ for the ice phase.

Equation 19.3 can now be written as

$$\eta = \frac{\pi^5}{\lambda^4} |K|^2 \sum_{i=1}^{N} D_i^6 \tag{19.6}$$

and we define the radar reflectivity factor Z as

$$Z = \sum_{i=1}^{N} D_i^6 \tag{19.7}$$

In radar meteorology, it is common to use the dimensions of millimeters for drop diameter D_i and to consider the summation to take place over a unit volume of size 1 m^3 to yield a volume density expression. Therefore, the conventional unit of Z is in mm^6/m^3. For ice particles, D_i is sometimes expressed as the diameter of the water droplet that would result if the ice particle were to melt completely. However, the radar scattering process for the many shapes and temperatures of ice particles is extremely complicated and a definitive generalized expression cannot be given.

It is often convenient to treat the drop or particle size distribution as a continuous function with a number density $N(D)$, where $N(D)$ is the number of drops per unit volume having diameters between D and $D + dD$. In this case, Z is given by the sixth moment of the particle size distribution,

$$Z = \int_0^\infty n(D) D^6 dD \tag{19.8}$$

If the radar beam is filled with scatterers, the sample volume of V is given[10] approximately by

$$V \approx \frac{\pi \theta \phi r^2 c \tau}{8} \tag{19.9}$$

where θ and ϕ are the azimuth and elevation beamwidths, c is the velocity of light, and τ is the radar pulse width. Substituting Eqs. 19.3 and 19.9 into Eq. 19.2, we see that the RCS for the distributed weather scatterer is directly proportional to the pulse volume as determined by the pulse length and antenna beam at the target range.

Then, combining Eqs. 19.2, 19.6, and 19.9 and substituting into Eq. 19.1 gives

$$P_r = \frac{\beta}{r^4}\frac{\pi\theta\phi r^2 c\tau}{8}\frac{\pi^5}{\lambda^4}|K|^2\sum_{i=1}^{N} D_i^6$$
$$= \frac{\beta\pi^6\theta\phi c\tau|K|^2}{8\lambda^4 r^2} Z \qquad (19.10)$$
$$= \frac{\beta' Z}{r^2}$$

This expression illustrates that for the distributed weather target the received power is (1) a function only of β' (a constant dependent upon all the radar system and physical parameters), (2) directly proportional to the radar reflectivity factor Z, and most significant, (3) inversely proportional to r^2 (not r^4 as in the case of point targets).

The radar system parameters included in β in Eq. 19.1 include the peak transmit power P_t, the antenna system gain G twice (once for transmitting and once for receiving), and the wavelength λ. We include all antenna system losses in this antenna system gain factor (radome, waveguide, rotary joints, etc.) since all the measurements must be referenced to the same point in the radar system—usually at a coupler near the circulator. Because the antenna gain is not uniform over the beamwidth, assuming a uniform gain can lead to errors in the calculation of Z. Using a similar derivation, Probert-Jones[34] took this into account, assumed a gaussian shape for the antenna beam, and derived the following equation for the received power:

$$P_r = \frac{P_t G^2\lambda^2\theta\phi c\tau}{512(2\ln 2)\pi^2 r^2}\sum_{i=1}^{N}\sigma_i \qquad (19.11)$$

where 2ln2 is the correction for the gaussian-shaped beam. By substituting Eqs. 19.3, 19.6, and 19.7 into Eq. 19.11, the received power can be expressed in terms of the reflectivity factor Z and range r as

$$P_r = \frac{P_t G^2\theta\phi c\tau\pi^3|K|^2 Z}{512(2\ln 2)\lambda^2 r^2} \qquad (19.12)$$

Because the receiving filter suppresses some of the received signal power, P_r must be reduced by L_r, which depends on the details of the transmitted spectrum and the receiver filter but is usually a factor of about 1.6 (2 dB) for a typical waveform and "matched filter." Solving for the radar reflectivity factor Z gives

$$Z = [1024\ln 2\ \lambda^2 L_r/P_t G^2\ \theta\phi c\ \tau\pi^3|K|^2]\ P_r r^2 \qquad (19.13)$$

where the reflectivity factor is expressed in terms of the received power and range.

One must be careful to use consistent units in Eq. 19.13. If meter-kilogram-seconds (mks) units are used, the calculation of Z will have dimensions of m^6/m^3. Conversion to the more convenient units of mm^6/m^3 requires that Z be multiplied by the factor 10^{18}. Furthermore, expressing P_r and r in common units of dBm (dB relative to 1 milliwatt) and km requires that Z also be multiplied by 10^3. Because Z values of interest can range over several orders of magnitude, a logarithmic scale is often used. Thus,

$$\text{dB}Z = C + P_r\,(\text{dBm}) + 20\log r\,(\text{km}) \qquad (19.14)$$

where C (isolated in brackets in Eq. 19.13) is the so called Weather Radar Constant with P_r expressed in dBm and r in km. Typical values of C are 65 to 75 dB for the operational weather radars. It is clear that for fixed range and received power, a lower value of the radar constant C allows a smaller reflectivity value in dBZ to be observed. Thus, smaller values of C correspond to more sensitive radars.

This equation can be used to measure the reflectivity factor Z when the antenna beam is filled, when the small scattering particle Rayleigh approximation is valid, and when the scatterers are in either the ice or the water phase. Because all these conditions are not always satisfied, it is common to use the term Z_e, the effective reflectivity factor, in place of Z. When Z_e is used, it is generally understood that the above conditions are assumed. Practitioners in the field of radar meteorology often use Z_e and Z interchangeably, albeit incorrectly.

Another factor we have ignored in the derivation of the radar equation is attenuation by precipitation and atmospheric gasses. At 10 cm wavelengths the attenuation is usually not significant; however, at 5, 3, and 2 cm and especially the yet shorter mm wavelengths, atmospheric attenuation must be accounted for in the radar equation by adding an additional range dependent term L_a. The following section gives details on estimating this attenuation in commonly encountered conditions.

Finally, it is important to note the Z values are of meteorological significance because they are directly related to cloud properties and actual rainfall rates R as described later in this chapter. Z values in nonprecipitating clouds as small as –40 dBZ are of interest for cloud physics studies. In the optically clear, lower atmospheric boundary layer, "clear air" Z values of the order –20 dBZ to 20 dBZ are typical and frequently originate from insects and birds.[35,36] In rain Z may range from about 0–10 dBZ to as much as 60 dBZ, with a 55 to 60 dBZ rain being of the type that can cause severe flooding. Severe hailstorms may produce Z values higher than 70 dBZ. Many operational radar types are designed to detect those Z values that produce measurable precipitation (0 to 60 dBZ) and "clear air" echoes to 100 km where the Earth's curvature prevents surface-based measurements. Thus, being able to measure strong precipitation echoes at short range and also weak precipitation echoes at long range requires radar receivers having a total dynamic range of 90–95 dB whereas measuring weak echoes in the presence of strong ground clutter requires as large an instantaneous dynamic range (> 60 dB) as possible. More recent operational radars and most research radars attempt to achieve the most sensitivity possible and can detect minimum reflectivity values of –40 dBZ or less at short ranges (e.g., 1 km).

Operational radars in the past have employed sensitivity time control (STC) to reduce gain at short range and compensate for strong nearby echoes; however, recent radars tend not to use STC techniques since receiver dynamic ranges are adequate to cover important weather echo intensities at the necessary ranges. Research radars have rarely used STC owing to the attendant loss of sensitivity at short ranges.

19.3 DESIGN CONSIDERATIONS

Four of the more significant factors that affect the design of meteorological radars are attenuation, range ambiguities, velocity ambiguities, and ground or sea clutter. The combination of these factors, along with the need to obtain adequate spatial resolution, leads to a wavelength selection in the range of 3 to 10 cm for most precipitation based applications.

Attenuation Effects. Attenuation has at least two negative effects on meteorological radar signals. First, making accurate quantitative measurements of the backscattered energy from precipitation at ranges farther than any intervening precipitation becomes very difficult. This inability to precisely measure the true backscattering cross section requires that quantitative measurements of precipitation rates be corrected for attenuation when possible.

Second, if the attenuation due to precipitation or the intervening medium is sufficiently great, the signal from a precipitation cell behind a region of strong absorption may be totally suppressed. One example of the potentially serious consequences of very strong absorption is the impact it might have on aviation storm avoidance radars, most of which are in the 3-cm band. It is common for short wavelength onboard aviation weather radars to not detect intense convective cells behind closer, high attenuation thunderstorms. Severe storms with high precipitation rates also cause high attenuation even at 5-cm wavelengths, as noted by Hildebrand[37] and Allen et al.[38]

In some meteorological radar applications, it is desirable to attempt to measure attenuation along selected propagation paths. This is done because absorption is related to liquid-water content and can provide useful information for the detection of such phenomena as hail, in accordance with the dual-wavelength technique described by Eccles and Atlas[39] and Vivek et al.[40]

In the following subsections, quantitative expressions relating attenuation to precipitation are given. Much of this is taken from Bean, Dutton, and Warner[22] and Lhermitte.[25] Battan[20] and Oguchi[41] are also excellent sources for additional information on the absorbing properties of precipitation.

Attenuation by Water Vapor. Atmospheric water vapor may take on values up to 25 g/m^3 and give variable attenuation depending on the water vapor content. However, at typical weather radar wavelengths longer than 3 cm, the attenuation is less than a few hundredths dB/km and is usually ignored. Gaseous oxygen contributes only a minor absorption effect at these centimeter wavelengths and is also usually ignored.

Attenuation in Clouds. Cloud droplets are regarded here as those water or ice particles having radii smaller than 100 μm (0.01 cm). For wavelengths of incident radiation well in excess of 0.5 cm, the attenuation depends primarily on the liquid-water content and is independent of the drop-size distribution. The generally accepted equations for attenuation by clouds usually show the moisture component of the equations in the form of the liquid-water content (grams per cubic meter). Observations[42] indicate that the liquid-water concentration in clouds generally ranges from 1 to 2.5 g/m^3, although Weickmann and Kampe[43] have reported isolated instances of cumulus congestus clouds (high towering convective clouds that frequently produce heavy precipitation) with water contents of 4.0 g/m^3 in the upper levels. In ice clouds, the water content rarely exceeds 0.5 and is often less than 0.1 g/m^3. The attenuation due to cloud droplets may be written

$$K = K_1 M \tag{19.15}$$

where K = attenuation, dB/km
K_1 = attenuation coefficient, dB/(km · g/m^3)
M = liquid-water content, g/m^3

$$M = \frac{4\pi\rho}{3} \sum_{i=1}^{N} a_i^3 \tag{19.16}$$

$$K_1 = 0.4343 \frac{6\pi}{\lambda} \text{Im}\left(-\frac{m^2 - 1}{m^2 + 2}\right) \tag{19.17}$$

where the a_i are droplet radii, ρ is the density of water, and Im is the imaginary part. Values of K_1 for ice and water clouds are given for various wavelengths and temperatures by Gunn and East[44] in Table 19.1.

Several important facts are demonstrated by Table 19.1. The decrease in attenuation with increasing wavelength is clearly shown. The values change by about an order of magnitude for a change of λ from 1 to 3 cm. The data presented here also shows that attenuation in water clouds increases with decreasing temperature. Ice clouds give attenuations about two orders of magnitude smaller than water clouds of the same water content. The attenuation of microwaves by ice clouds can be neglected for practical purposes.

Attenuation by Rain. Ryde and Ryde[45] calculated the effects of rain on microwave propagation and showed that absorption and scattering effects of raindrops become more pronounced at the higher microwave frequencies, where the wavelengths and the raindrop diameters are more nearly comparable. In the 10-cm band and at shorter wavelengths, the effects are appreciable, but at wavelengths in excess of 10 cm, the effects are greatly decreased. It is also clear that suspended water (cloud) droplets and rain have an absorption rate in excess of that of the combined oxygen and water-vapor absorption.[46]

In practice, it has been convenient to express rain attenuation as a function of the precipitation rate R, which depends on the liquid-water content and the fall velocity of the drops, the latter in turn depending on the size of the drops. Ryde[47] studied the attenuation of microwaves by rain and deduced, by using Laws and Parsons's[48] drop size distributions, that this attenuation in decibels per kilometer can be approximated by

$$K_R = \int_0^{r_0} K[R(r)]^{\alpha} dr \tag{19.18}$$

where K_R = total attenuation, dB
$R(r)$ = rainfall rate along path r
r_0 = length of propagation path, km
K = constant dependent on frequency and temperature
α = constant dependent on frequency

TABLE 19.1 One-way Attenuation Coefficient K_1 in Clouds in dB/km per g/m^3 Liquid Water[44]

		Wavelength, cm			
Temperature, °C		0.9	1.24	1.8	3.2
Water cloud	20	0.647	0.311	0.128	0.0483
	10	0.681	0.406	0.179	0.0630
	0	0.99	0.532	0.267	0.0858
	–8	1.25	0.684	0.34$^+$	0.112$^+$
Ice cloud	0	8.74×10^{-3}	6.35×10^{-3}	4.36×10^{-3}	2.46×10^{-3}
	–10	2.93×10^{-3}	2.11×10^{-3}	1.46×10^{-3}	8.19×10^{-3}
	–20	2.0×10^{-3}	1.45×10^{-3}	1.0×10^{-3}	5.63×10^{-3}

$^+$ extrapolated value

Medhurst[49] shows that $\alpha = 1$ is a good assumption in many cases. The path loss per mile for the three frequency bands of 4, 6, and 11 GHz is shown in Figure 19.1. Lhermitte[25] extends this early work to higher frequencies, confirming the relationship and describing the effect of the drop-size distribution on the attenuation rate. He also reviews more recent empirical data.

The greatest uncertainty in prediction of attenuation rates caused by rainfall, when theoretical formulas are used as a basis for calculation, is the extremely limited knowledge of drop-size distributions in rain of varying fall rates under differing climatic and weather conditions. Lhermitte[25] and Uijlenhoet et al.[50] thoroughly review the evolution of analytic expressions for drop-size distributions, describing the Marshall-Palmer experiment and their resulting exponential distribution and the more general three-parameter Gamma distribution. They also review the dependence of the parameters defining the distribution to the rain rate and to the type of rainfall. There is little evidence that rain with a known rate of fall has a unique drop-size distribution, although Burrows and Attwood's studies seem to indicate that a certain most probable drop-size distribution can be attached to rain of a given rate of fall.[51] Results of this study are shown in Table 19.2, which gives the percentage of total volume of rainfall occupied by raindrops of different diameters and varying rainfall rates (millimeters per hour). In the basis of these results, the absorption cross section of different rain rates is shown in Table 19.3. This table gives the decibel attenuation per kilometer for different rainfall rates for radar wavelengths between 0.3 and 10 cm.

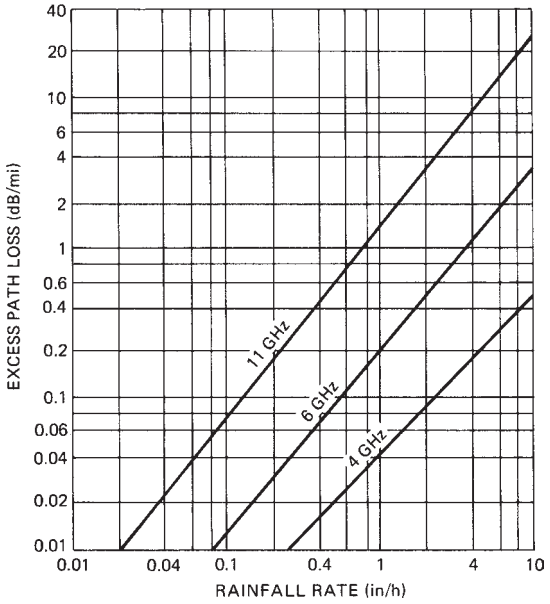


FIGURE 19.1 Theoretical rain attenuation in dB/mile (statute) versus rainfall rate (*after J. W. Ryde and D. Ryde*[45])

TABLE 19.2 Drop Size Distributions at Different Precipitation Rates[51]

	Precipitation Rate R, mm/h							
	0.25	1.25	2.5	12.5	25	50	100	150
Drop Diameter D, cm	Percentage of a Given Volume Containing Drops of Diameter D							
0.05	28.0	10.9	7.3	2.6	1.7	1.2	1.0	1.0
0.10	50.1	37.1	27.8	11.5	7.6	5.4	4.6	4.1
0.15	18.2	31.3	32.8	24.5	18.4	12.5	8.8	7.6
0.20	3.0	13.5	19.0	25.4	23.9	19.9	13.9	11.7
0.25	0.7	4.9	7.9	17.3	19.9	20.9	17.1	13.9
0.30	.....	1.5	3.3	10.1	12.8	15.6	18.4	17.7
0.35	.....	0.6	1.1	4.3	8.2	10.9	15.0	16.1
0.40	.....	0.2	0.6	2.3	3.5	6.7	9.0	11.9
0.45	.....	.....	0.2	1.2	2.1	3.3	5.8	7.7
0.50	.....	.....	.....	0.6	1.1	1.8	3.0	3.6
0.55	.....	.....	.....	0.2	0.5	1.1	1.7	2.2
0.60	.....	.....	.....	.....	0.2	0.5	1.0	1.2
0.65	.....	.....	.....	.....	.....	0.2	0.7	1.0
0.70	.....	.....	.....	.....	.....	.....	.....	0.3

Since the total-attenuation cross section[52] depends on the temperature (because of its effects on the dielectric properties of water), it is important to evaluate the attenuation of rains whose drops are at different temperatures from those in the preceding tables. Table 19.4 contains the necessary data relative to the change of attenuation with temperature and can be used with Table 19.3.

To determine total attenuation caused by rainfall through a particular precipitation path, something must be known or assumed about the nature of the precipitation itself and, consequently, about how its rainfall rates and drop sizes are distributed in three dimensions.

A systematic vertical variation of R, decaying with height above a measured surface value, seems to be appropriate in *stratiform*[53] rainfall, which is rain having a widespread and continuous nature. Such widespread rainfall is usually triggered by

TABLE 19.3 Attenuation in Decibels per Kilometer for Different Rates of Rain Precipitation at a Temperature of 18°C Using the Drop-size Distributions in Table 19.2[51]

Precipitation	Wavelength λ, cm								
Rate R, mm/h	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 1.0$	$\lambda = 1.25$	$\lambda = 3.0$	$\lambda = 3.2$	$\lambda = 10$
0.25	0.305	0.230	0.160	0.106	0.037	0.0215	0.00224	0.0019	0.0000997
1.25	1.15	0.929	0.720	0.549	0.228	0.136	0.0161	0.0117	0.000416
2.5	1.98	1.66	1.34	1.08	0.492	0.298	0.0388	0.0317	0.000785
12.5	6.72	6.04	5.36	4.72	2.73	1.77	0.285	0.238	0.00364
25.0	11.3	10.4	9.49	8.59	5.47	3.72	0.656	0.555	0.00728
50	19.2	17.9	16.6	15.3	10.7	7.67	1.46	1.26	0.0149
100	33.3	31.1	29.0	27.0	20.0	15.3	3.24	2.80	0.0311
150	46.0	43.7	40.5	37.9	28.8	22.8	4.97	4.39	0.0481

TABLE 19.4 Correction Factor (Multiplicative) for Rainfall Attenuation[51]

Precipitation Rate R, mm/h	λ, cm	0°C	10°C	18°C	30°C	40°C
0.25	0.5	0.85	0.95	1.0	1.02	0.99
	1.25	0.95	1.00	1.0	0.90	0.81
	3.2	1.21	1.10	1.0	0.79	0.55
	10.0	2.01	1.40	1.0	0.70	0.59
2.5	0.5	0.87	0.95	1.0	1.03	1.01
	1.25	0.85	0.99	1.0	0.92	0.80
	3.2	0.82	1.01	1.0	0.82	0.64
	10.0	2.02	1.40	1.0	0.70	0.59
12.5	0.5	0.90	0.96	1.0	1.02	1.00
	1.25	0.83	0.96	1.0	0.93	0.81
	3.2	0.64	0.88	1.0	0.90	0.70
	10.0	2.03	1.40	1.0	0.70	0.59
50.0	0.5	0.94	0.98	1.0	1.01	1.00
	1.25	0.84	0.95	1.0	0.95	0.83
	3.2	0.62	0.87	1.0	0.99	0.81
	10.0	2.01	1.40	1.0	0.70	0.58
150	0.5	0.96	0.98	1.0	1.01	1.00
	1.25	0.86	0.96	1.0	0.97	0.87
	3.2	0.66	0.88	1.0	1.03	0.89
	10.0	2.00	1.40	1.0	0.70	0.58

a relatively large-scale mechanism, such as a frontal or monsoon situation. A vertical variation of R of the form

$$R = R_0 e^{-dh^2} \tag{19.19}$$

can be assumed to be appropriate under continuous rainfall conditions.[54] R_0 is the surface rainfall rate, h is the height above the Earth's surface, and d is a constant, equal to about 0.2. Convective precipitation, however, shows a quite different nature. For example, the presence of *virga*[53] (precipitation aloft but evaporating before reaching the surface) associated with so many shower-type clouds in dry climates indicates that convective shower rainfall profiles are much more difficult to model.

Attenuation by Hail. Ryde[47] concluded that the attenuation caused by hail is one-hundredth of that caused by rain and that ice-crystal clouds cause no sensible attenuation and show very small attenuation even at the excessive rate of fall of 5 in/h. However, the scattering by ice spheres surrounded by a concentric film of liquid water having a different dielectric constant does not give the same effect that Ryde's results for dry particles would indicate.[55] For example, when one-tenth of the radius of an ice sphere of radius 0.2 cm melts, scattering of 10-cm radiation is approximately 90% of the value that would be scattered by an all-water drop.

At wavelengths of 1 and 3 cm with $2a = 0.126$ (a = radius of drop), Kerker, Langleben, and Gunn[55] found that particles attained total attenuation cross sections corresponding to all melted particles when less than 10% of the ice particles was melted. When the melted

mass reached about 10 to 20%, the attenuation was about twice that of a completely melted particle. These calculations show that the attenuation in the melting of ice immediately under the 0°C isotherm can be substantially greater than in the snow region just above and, under some circumstances, greater than in the rain below the melting level. Further melting cannot lead to much further enhancement, apparently, and may lead to a lessening of the reflectivity of the particle by bringing it to sphericity or by breaking up the particle. Melting of ice particles produces enhanced backscatter, and this effect gives rise to the observed elevated *bright band*[53] near the 0°C isotherm.

Lhermitte[25] discusses hail attenuation for shorter wavelength radars when resonant region (Mie) scattering is the dominant scattering mechanism. Using accepted size distributions of dry hail, he shows attenuation rates over the frequency interval of 3–150 GHz that are negligible at the lower frequencies but rise asymptotically to about 3 dB/km at frequencies above 100 GHz.

Attenuation by Fog. The characteristic feature of a fog is the reduction in visibility. *Visibility*[53] is defined as "the greatest distance in a given direction at which it is just possible to see and identify with the unaided eye (1) in the daytime, a prominent dark object against the sky at the horizon and (2) at night a known, preferably unfocused, moderately intense light source." Although the visibility depends upon both drop size and number of drops and not entirely upon the liquid-water content, in practice the visibility is an approximation of the liquid-water content and, therefore, may be used to estimate radio-wave attenuation.[56]

On the basis of Ryde's work, Saxton and Hopkins[57] give the figures in Table 19.5 for the attenuation in a fog or clouds at 0°C temperature. The attenuation varies with the temperature because the dielectric constant of water varies with temperature; therefore, at 15°C and 25°C the figures in Table 19.5 should be multiplied by 0.6 and 0.4, respectively. It is immediately noted that cloud or fog attenuation is an order of magnitude greater at 3.2 cm than at 10 cm and that nearly another order of magnitude increase occurs between 3.2 and 1.25 cm.

Range and Velocity Ambiguities. Weather radars utilize a sequence of pulses to measure the radar reflectivity and doppler characteristics. Because the pulsing rate typically determines the sampling frequency for doppler quantities on a constant pulsing rate radar, the unambiguous doppler frequency (Nyquist frequency) for a fixed pulse-repetition-frequency (PRF) radar is given by

$$F_{\text{Nyq}} = \pm \text{PRF}/2 \tag{19.20}$$

where PRF is the pulse repetition frequency. Simultaneously, the unambiguous range interval is given by

$$R_a = \frac{c}{2 \cdot \text{PRF}} \tag{19.21}$$

TABLE 19.5 Attenuation Caused by Clouds or Fog Temperature = 0°C[57]

	Attenuation, dB/km		
Visibility, m	$\lambda = 1.25$ cm	$\lambda = 3.2$ cm	$\lambda = 10$ cm
30	1.25	0.20	0.02
90	0.25	0.04	0.004
300	0.045	0.007	0.001

and the product $F_{\text{Nyq}} R_a$ is simply

$$F_{Nyq} R_a = \frac{c}{4} \tag{19.22}$$

Since the doppler shift f and the target radial velocity v are linearly related, the unambiguous velocity is related to the Nyquist frequency by

$$V_a = \frac{\lambda}{2} F_{\text{Nyq}} \tag{19.23}$$

It follows that the product of unambiguous velocity and unambiguous range is

$$V_a R_a = \frac{\lambda c}{8} \tag{19.24}$$

For constant PRF radars, this product is maximized by maximizing λ, the transmitted wavelength. Thus, using longer wavelengths allows optimizing the PRF by trading unambiguous range for unambiguous velocity. For standard constant PRF radars, 10-cm wavelengths have been widely chosen for most precipitation measurements of interest in which design parameters such as radar beamwidth, size of antenna, and attenuation effects may be made acceptable.

Ground Clutter Effects. Many meteorological radar applications call for the detection of precipitation echoes in the presence of ground clutter. Specifically, precipitation measurements near the ground are of extreme interest in agricultural and hydrological applications as well as information for the general public. Applications in which ground clutter is serious relates to the ground-based radar detection of low-level wind shear at airports and measuring precipitation near the ground in mountainous terrain for flash flood warnings. Both the Nexrad and Terminal Doppler Weather Radar (TDWR) network radars have been designed to assure clutter suppression in excess of 40 dB.[7,58]

Although ground clutter cannot be eliminated, its effect can be mitigated through careful design. The primary approach is to use an antenna with low sidelobes, particularly in elevation, which will suppress the clutter component of the input echo when the main beam is slightly above the horizon. A second approach is using shorter wavelengths. Shorter wavelengths result in improved signal-to-clutter ratios owing to the fact that the Rayleigh scattered weather signal power is inversely proportional to λ^4, whereas the ground clutter return is only weakly dependent on wavelength. If one assumes that the clutter signal is wavelength independent and the antenna beamwidth is fixed, it can be shown that the weather-signal-power to clutter-power ratio is inversely proportional to λ^2.

Weather radars typically use digital signal processing techniques to implement clutter filters that suppress near zero velocity clutter echoes.[59] These filters may be implemented using either a time domain clutter filter applied to the I and Q radar video data (one form of which is sometimes called a *delay line canceler* from its early analog implementation) to suppress the zero velocity ground clutter components or a frequency domain doppler power spectrum (a digital "filter bank") to achieve the same effect.[32] The time domain filters for mechanically scanned weather radars are usually infinite impulse response (IIR) filters with narrow, but adjustable, widths up to a few m/s and having suppression levels of 40–60 dB and very steep transition regions.[60] These time domain filters with a frequency notch centered at zero velocity (frequency)

will also suppress weather echo power that may exist in the same velocity region and bias all the estimates of reflectivity, velocity, and width.

Spectral domain clutter filters implemented by a discrete Fourier transform (DFT), on the other hand, suppress the near zero clutter components in the frequency domain and may interpolate the remaining spectrum across this region to retain most of the underlying (signal or noise) spectral information. An alternative frequency domain technique for the Nexrad radar separately models the clutter and weather signal as gaussian-shaped spectra and separates these two components of the doppler spectrum using digital search algorithms and then removes these clutter components while leaving the underlying weather signal unperturbed.[61] Thus, when the gaussian assumptions apply, the remaining weather signal spectrum provides an unbiased estimator for all the weather parameter spectrum moment estimates.

Typical Weather Radar Designs. There is no universal weather radar system design that can serve all purposes. Airborne weather radars are constrained by size and weight limitations. Ground-based radars may be constrained by cost and siting considerations. Severe storm warning radars require long range and high unambiguous velocity and then must penetrate very heavy rain, thus dictating long wavelengths. Radars designed for studies of nonprecipitating clouds typically use short wavelengths[62,63] (8 mm and 3 mm) in order to achieve sufficient sensitivity to detect small cloud particles of the order of 10–100 μm at sufficiently small resolution volumes. Sensitive, short-range FM-CW radars[64] having high average power can be used to obtain very high range resolution for detecting very thin scattering layers in the clear air boundary layer.

Most meteorological radars are pulsed radars having doppler capability. Ground-based radars used for severe storm research and warnings normally use S-band (~3 GHz) or C-band (~5.5 GHz) transmitters. Airborne weather avoidance and precipitation radars primarily use X band (~10 GHz) due to size limitations and occasionally C-band transmitters to minimize attenuation. Airborne and ground cloud radars and spaceborne radars encompass the mm-wavelengths at K_u band (~15 GHz), K_a band (~35 GHz), and W band (~94 GHz).

Beamwidths of ≤ 1° are commonly used for longer-range radars. Admittedly, this is somewhat arbitrary, but the choice of 1° is based upon several decades of experience. A 1° beam will provide a cross-range resolution of 1 km at a range of 60 km. Because thunderstorms contain important spatial features, such as heavy precipitation shafts and updraft cores, with horizontal dimensions of the order 1 to 5 km, a 1° beam is reasonably well matched to these atmospheric phenomena being observed out to ranges of a few hundred km. Shorter-range airborne weather radars often employ beamwidths of 2–3° as a compromise between wavelength requirements and antenna size constraints whereas spaceborne radars may use a fraction of a degree beamwidth to retain usable horizontal resolution at typical long ranges (250–500 km).

Operational weather radars normally are capable of short and long pulse operation in the range of 0.5 μs to about 6 μs and PRFs between 300 and 3000 Hz for long-range precipitation radars. Through pulse-width diversity, high resolution is obtained (usually at short range) whereas for long-range detection, longer pulses provide increased sensitivity and tend to equalize the along-beam and cross-beam resolutions. The shorter wavelength K_u-, K_a-, and W-band radars typically use pulse lengths less than 1 μs to achieve improved range resolution and PRFs between 3000 and 10,000 Hz because of the short-range cloud measurements that are limited by attenuation

at these wavelengths. Space-borne radars also use these higher PRFs but keep track of multiple pulses en-route to their weather regions far below the orbital altitudes.

Equation 19.12 shows that the received power is directly proportional to the pulse length τ. The noise power P_n is conventionally given by

$$P_n = \kappa TB \tag{19.25}$$

where κ = Boltzmann's constant, 1.38×10^{-23} W/(Hz · K)
T = receiver noise temperature, K
B = receiver noise bandwidth, Hz

For a receiver filter matched to the pulse length,

$$B \approx \frac{1}{\tau} \tag{19.26}$$

Sometimes weather radars will use a short pulse for high PRF doppler processing at short ranges and a longer pulse at low PRF for greater sensitivity when performing long-range surveillance scans to monitor distant weather. Since the transmitted peak power is typically constrained to be fixed, then the transmitted average power increases linearly with τ. Also, the matched filter bandwidth and associated noise power decreases inversely with τ. If the radar pulse volume is filled with distributed meteorological scatterers, then the radar cross section of the weather target also increases with τ (as determined by Equations 19.9 and 19.12) and the signal-to-noise ratio (SNR) of received power is proportional to τ^2:

$$\frac{P_r}{P_n} \propto \frac{\tau}{\kappa TB} \approx \frac{\tau^2}{\kappa T} \tag{19.27}$$

Thus, under these common conditions increasing the pulse length will increase the SNR and the effective radar range. It is important to note that the distributed target radar dependence of SNR on τ^2 is different from the point target radar in which the matched filter SNR equals the ratio of pulse energy to noise spectral density ($2E/N_0 = 2P_t\ \tau/N_0$), a linear dependence on τ. The squared dependence for distributed scatterers is because the transmit pulse scatters power from all scatterers in the $c\tau/2$ pulse volume (not just a single point target) thereby increasing the radar cross section of the weather scatterer.

PRFs for meteorological radars range from as low as several hundred s^{-1} for long-range detection to several thousand s^{-1} for shorter-wavelength systems attempting to achieve high unambiguous velocities. Generally speaking, most meteorological doppler radars are operated in a single PRF mode, compromising the radar's ability to unambiguously resolve either range or velocity. However, the pulsing sequence may use a "dual PRF" mode in which groups of constant PRF pulses are transmitted or a "dual (staggered) PRT" (pulse repetition time) to resolve both range and velocity ambiguities.[65] Another approach is to employ a transmitted pulse sequence with random[66] or deterministic[67] phases from pulse to pulse, which allows overlaid echoes to be separated. Multiple PRT techniques have also been explored but are not in common use.[68] Range ambiguities cannot be totally eliminated, but their effects can be significantly mitigated through these approaches.

To discuss design details of the common types of meteorological radars is beyond the scope of this chapter. Rinehart[26] gives a detailed table of system characteristics of

TABLE 19.6 Relevant Nexrad System Characteristics[69]

Transmitted Peak/Average Power (klystron)	750 kW / 1500 W
Pulse length	225, 675 m (1.57, 4.50 μs)
Polarization	Linear horizontal
Wavelength	10.6 cm
Receiver noise temperature	450 K
Dynamic range	95 dB
Antenna gain	45.5 dB
Beamwidth	0.95°
Sidelobe levels	< –27 dB
Maximum range (reflectivity data)	460 km
Maximum range (doppler data)	230 km
Unambiguous velocity	± 50 m/s
Clutter suppression (maximum)	55 dB
System sensitivity	–7.5 dBZ at 50 km
Rotation rate	10–30 deg/sec

a variety of weather radars. However, it will be useful herein to include some of the important characteristics of the Nexrad WSR-88D radar, which illustrate the performance of modern operational weather radars. Table 19.6 contains some of the more relevant original Nexrad design features.

Figure 19.2 shows a typical Nexrad installation at Missoula, Montana. The antenna is mounted on a tower to clear surrounding obstacles such as buildings and trees. The electronic equipment is housed in one shelter and the standby generator is housed in another.

FIGURE 19.2 Nexrad WSR-88D radar at Missoula, Montana, mounted on 15-m tower with two equipment shelters—one containing the transmitter, receiver, processor, and communications equipment, and the other containing a standby generator (*Photo courtesy of NOAA/NWS*)

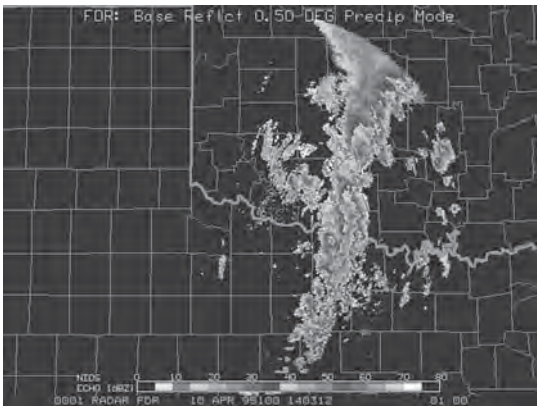


FIGURE 19.3 Nexrad reflectivity data from Frederick, Oklahoma, radar on April 10, 1995, showing line of intense convective cells and surrounding precipitation (*Photo courtesy of NOAA/NWS*)

Figure 19.3 shows an example of radar reflectivity data from the Frederic, Oklahoma, radar as a line of intense thunderstorms and associated rainfall pass through the coverage area. The processing techniques used to generate the various weather images and products are discussed later in this chapter.

Figure 19.4 shows the S-band (10 cm) and K_a-band (0.8 cm) dual-polarization doppler research radar[70] operated by the National Center for Atmospheric Research (NCAR). The system permits simultaneous measurements of the reflectivity factor on two wavelengths, doppler parameters on the S-band wavelength, and extensive polarimetric measurements at both wavelengths. The technical characteristics are similar to the Nexrad specifications. The antenna beams are matched with approximately 1° beamwidth and the large S-band dish is 8.7 m in diameter. The peak transmitted power

FIGURE 19.4 S-Pol: The multi-parameter S-band and K_a-band polarimetric research radar pointing at the sun for a solar calibration at the National Center for Atmospheric Research, Boulder, Colorado (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

is 1 MW at S band and 50 KW at K_a band. The pulse widths are approximately 1 μs. The PRF is typically 1000 s^{-1} at S band and a few times larger at K_a band. This radar system is characteristic of the technologies currently in place in the research community.

Polarimetric Radar. Meteorological radars using dual polarization transmit and receive both horizontal and vertical polarizations to estimate additional characteristics of the weather targets.[13,14,24] Transmitting the two orthogonal polarizations either simultaneously (SHV) or transmitting them separately in a predetermined sequence and using dual parallel digital receivers (one on each polarization channel) allow estimation of the differential quantities between echoes from the two polarizations. One can derive improved rainfall rates (as well as other physical information on the type of precipitation) as functions of polarimetric measurements that relate the differences in the horizontally and vertically polarized signals. By far the most common polarimetric parameters are the "differential reflectivity" (Z_{dr}) and "differential phase" (Φ_{dp}), which give bulk scattering and propagation characteristics of the meteorological targets. Letting E_h and E_v denote the complex received signal *voltages* that also represent the received *electric fields* at horizontal and vertical polarizations, the important polarimetric parameters to be estimated are given below[11]:

$$\text{Differential reflectivity } Z_{dr} = \langle Z_h \rangle / \langle Z_v \rangle \tag{19.28}$$

$$\text{Differential phase } \Phi_{dp} = \langle \Phi_v \rangle - \langle \Phi_h \rangle \tag{19.29}$$

$$\text{Specific differential phase } K_{dp} = d\langle \Phi_{dp} \rangle / \text{dr} \tag{19.30}$$

$$\text{Co-polar correlation ratio } \rho_{hv} = |\langle E_h^* E_v \rangle| / \langle E^*_h E_h \rangle^{1/2} \langle E^*_v E_v \rangle^{1/2} \tag{19.31}$$

$$\text{Linear depolarization ratio LDR} = \langle Z_{cx\,v} \rangle / \langle Z_{co\,h} \rangle \tag{19.32}$$

Z_h and Z_v are the measured reflectivities of the horizontal and vertical co-polarized received signals and Z_{dr} is expressed in dB, whereas Φ_h and Φ_v are the measured phases of the same polarized received signals. K_{dp} is the suitably smoothed range derivative of the measured differential phase, Φ_{dp}, and is usually expressed in deg/km. ρ_{hv} is the co-polar correlation coefficient of E_h and E_v where the phase measurements are assumed to be time coincident, which is the case with SHV transmission and reception. LDR is a ratio of the cross-polar vertical reflectivity ($Z_{cx\,v}$) normalized by the co-polar horizontal reflectivity ($Z_{co\,h}$). Because polarimetric measurements add new dimensions of radar information and because these measurements are related to the physical characteristics of the scatterers, suitable combinations of these data give a strong indication of precipitation type (rain, snow, ice particles, sleet, hail, etc)[16,71] as well as radar echo classification into various categories (precipitation, ground or sea clutter, birds and insects, chaff, etc).[72]

Radar Calibration. To effectively use radars for accurate precipitation estimates, the conversion of measured reflectivity factor to received echo power must be well known. The gains (or conversion constants) of the various radar components can be measured using engineering test equipment, manufacturer specifications, and field measurements. Calibrating weather radar usually means to accurately specify the radar constant in the radar equation and to accurately estimate the received power

measured by the radar system. It also encompasses items such as knowing where the scattering volume is in 3D space by knowing antenna pointing angles and accurately determining range.

Although floating spheres, tetrahedral reflectors, and other targets of known radar cross section may be used,[26,73] the important "solar calibration" technique uses the solar position to adjust antenna pointing and the radiated solar flux to determine antenna gain.[74,75] Along with other radar parameter measurements, one can readily determine the radar constant. By injecting test signals of known power, one can determine the receiver gain transfer function. For polarimetric radars special care is required.[76] It has been demonstrated that by measuring the cross-polar powers in addition to the solar flux, an accurate estimate of the differential channel calibration can be obtained.[77] The AMS held a very successful Weather Radar Calibration Symposium[78] documenting all aspects of meteorological radar calibration.

19.4 SIGNAL PROCESSING

To compute the various weather products necessary for forecasting, warning, and other operational activities, the first three spectrum moments corresponding to received power, mean radial velocity, and doppler spectrum or velocity width must be estimated. Keeler and Passarelli[79] review standard estimation techniques and errors for spectrum moment estimation. For the highest resolution measurements, these spectral moments must be computed at each range gate sensed by the radar and converted into meaningful meteorological information. The nature of the distributed meteorological target, which shall be discussed next, imposes some special requirements not generally considered for hard, nonfluctuating targets on the processing techniques to estimate these moments.

It can be shown that the received signal from meteorological targets is well represented by a narrowband gaussian process.[23] This is a direct consequence of the fact that (1) the number of scatterers in the pulse volume is large ($>10^6$); (2) the pulse volume is large compared with the transmitted wavelength; (3) the pulse volume contains multiple point scattering sources, causing all phases on the range from 0 to 2π to be combined and returned; and (4) the particles are in relative motion due to turbulence, wind shear, and their varying fall speeds.

The superposition of the scattered electric fields from such a large number of particles (each with different amplitude and random phase) gives rise, through the central limit theorem, to a signal with bivariate, or a two dimensional, gaussian probability density function. Thus, the fluctuating amplitude of the return signal has a Raleigh statistical distribution and its phase is uniformly distributed between 0–2π. Furthermore, its intensity (power) is exponentially distributed.[80,81] Because all the particles within the sample volume are moving with some mean or average radial velocity, there is a mean frequency of the doppler spectrum that is shifted from the transmitted frequency. Finally, since the particles are in motion with respect to one another, there is also a doppler spread, often referred to as the width of the doppler spectrum. Zrnic describes a straightforward technique of synthesizing digital weather radar signals from a parameterized doppler spectrum characterizing a specific pulse volume.[82] Doviak and Zrnic[23] as well as Bringi and Chandrasekar[24] give detailed derivations of these relationships, whereas Keeler and Passarelli[79] summarize distributed target data characteristics and relate them to the sampled data sets representative of weather radar and other atmospheric sounding systems.

Spectrum Moment Estimation. A common Gaussian model of the mean received power spectral density of a meteorological signal[82] is depicted in Figure 19.5 and can be interpreted as follows. The received power is simply the integral under the curve (the zeroth moment) and is given by

$$P_r = \int S(f)df = \int S(v)dv \tag{19.33}$$

where f and v are related by $f = (2/\lambda)v$.

The mean velocity ($\bar{v}$) is given by the first moment of the spectrum:

$$\bar{v} = \frac{\int vS(v)dv}{\int S(v)dv} \tag{19.34}$$

The spectrum (velocity) width (σ_v) is given by taking the square root of the second central moment:

$$\sigma_v^2 = \frac{\int (v-\bar{v})^2 S(v)dv}{\int S(v)dv} \tag{19.35}$$

Radar meteorologists sometimes refer to σ_v^2 as the spectrum variance because of its computational equivalence to the variance of a continuously distributed random variable. In short, $S(v)$ is analogous to a probability density function for v since it is actually a reflectivity weighted distribution of particle velocities within the scattering volume. The term *spectrum width* will be used to refer to σ_v. It is clear, therefore, that the doppler spectrum contains the information necessary to measure meteorologically important signal parameters. These first three moments are usually referred to as *base data* and often labeled Z, V, and W with the appropriate conversions and units.

In the most general case, quadrature phase detection is used to obtain the real and imaginary parts of the complex signal. These are usually digitized in a large number of range gates (≈1000) at the radar's pulse repetition frequency. The resultant complex time series in each gate can then be processed by using a fast Fourier transform (FFT) to obtain an estimate of the doppler power spectrum[83] from which the echo power, mean velocity, and spectrum width can be obtained.

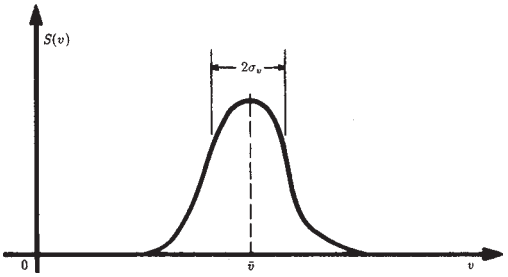


FIGURE 19.5 Gaussian model of the mean doppler power spectrum. The three spectral moments (received power, radial velocity, and spectrum width) can be estimated from the spectrum and are directly related to the meteorological variables of interest.

An efficient moment estimation technique was originally described by Rummler[84] and reinterpreted by Doviak and Zrnic.[23] This estimator makes use of the fact that the complex autocorrelation function of the received signal has the general form

$$R(nT) = P_r\ \rho(nT)\ \exp\left[j\frac{4\pi\bar{v}}{\lambda}nT\right] \tag{19.36}$$

where $\rho(nT)$ is the correlation coefficient of the time series data and nT is the time lag.

It can be shown that $\bar{v}$, the mean velocity, is a function of the first lag $R(T)$

$$\bar{v} = \frac{\lambda}{4\pi T}\arg[R(T)] \tag{19.37}$$

It can also be shown[85] that

$$\sigma_v^2 \approx \frac{\lambda^2}{8\pi^2T^2}\left[1-\frac{R(T)}{R(0)-P_n}\right] \tag{19.38}$$

where P_n is the noise power.

This estimator has been widely used in the past for mean-frequency estimation with doppler meteorological radars. The estimates are unbiased in the presence of noise when the doppler spectrum is symmetrical. Its greatest appeal, however, is due to its computational simplicity. For a pulsed radar, with a pulse repetition period T, $R(T)$ is obtained from adjacent pairs of pulses using the simple expression[86]

$$R(T) = \frac{1}{N}\sum_{k=0}^{N-1} s_{k+1}s_k^* \tag{19.39}$$

where the s_k are the complex in-phase (I) and quadrature (Q) signal samples (sampled at the radar PRT) at a given range and s_k^* is its complex conjugate. It is clear that this algorithm requires N complex multiplications for a time series of N samples whereas the FFT requires $N \log_2 N$. This so-called *pulse-pair algorithm*, therefore, is an efficient spectral moment estimation technique. However, it should be applied only when one is certain to have a pure weather signal in white noise; otherwise, other spectral components or non-white noise will bias the spectral moment estimates.

For many past operational radars, the pulse-pair processor was the technique of choice. However, in many research applications, it remains advantageous for improved data quality control purposes to access the full doppler spectrum and remove artifacts prior to computing the spectral moments corresponding to the meteorological data of interest.[79,87] The wind profiling radar community has made extensive use of spectral processing for artifact removal and sensitivity improvement.[88] One of the more important tasks is estimating the noise floor so as to remove its effect on the spectral moment estimates. Two objective techniques are common.[89,90] Constantly improving programmable digital signal processing chips and signal processing computers make it possible for radar meteorologists to implement various types of spectrum processing techniques that vastly improve data quality over the simple pulse pair processing algorithm. Furthermore, processors that adapt themselves to a variable environment in which they usually operate are feasible. Flexible programming of digital signal processors permits tailoring of the processor's characteristics to the application from day to day or even beam to beam and gate to gate.

Measurement Accuracy. Because the received signals are sample functions from gaussian random processes, the doppler spectrum and its moments cannot be measured exactly in any finite period of time. Consequently, all measurements will be somewhat in error, with the error being a function of the properties of the atmosphere, the radar wavelength, and the time allocated to the measurement.

The theoretical development of signal estimator statistics is found in Denenberg, Serafin, and Peach[91] for the FFT technique. Doviak and Zrnić[23] cover the subject quite completely while Keeler and Passarelli[79] provide a good summary of all these estimation techniques and the respective measurement error expressions. Following are some useful expressions for the mean square error of mean power and mean velocity estimates.

Power Estimation. It is well known[80] that for a gaussian random process using square-law signal detection, samples of the mean power P_r of the process are exponentially distributed with variance P_r^2. The variation is due to the process itself, not any noise associated with the measurement. Given a time T_0(sec) allocated to the measurement and a signal bandwidth σ_f(Hz), there will be approximately $\sigma_f T_0$ independent samples of the square of the signal envelope. It follows, therefore, that for the high SNR case an estimate of the mean power for this process $\hat{P}_r$ will have a variance or mean square error given by

$$\text{var}(\hat{P}_r) \approx \frac{P_r^2}{\sigma_f T_0} \tag{19.40}$$

Substituting for σ_f from the expression $\sigma_f = 2\sigma_v/\lambda$, where σ_v is the width of the doppler spectrum, Eq. 19.40 becomes

$$\text{var}(\hat{P}_r) \approx \frac{\lambda P_r^2}{2\sigma_v T_0} \tag{19.41}$$

This expression is valid for high signal-to-noise cases.

Velocity Estimation. The variance of the mean frequency estimate of the doppler spectrum is

$$\text{var}(\hat{f}) = \frac{1}{P_r^2 T_0} \int f^2 S^2(f + \bar{f}) df \tag{19.42}$$

This is an interesting result showing that the variance of the estimate $\hat{f}$ is a function only of the shape of the doppler spectrum (primarily its spectrum width) and the integration time T_0 allocated for processing. If the spectrum can be accurately modeled by a gaussian shape with variance σ_f^2, Eq. 19.42 reduces to

$$\text{var}(\hat{f}) = \frac{\sigma_f}{4\sqrt{\pi} T_0} \tag{19.43}$$

Noting that $\text{var}(\hat{v}) = (\lambda/2)^2 \text{var}(\hat{f})$, we can write

$$\text{var}(\hat{v}) = \frac{\lambda \sigma_v}{8\sqrt{\pi} T_0} \tag{19.44}$$

If we multiply the numerator and denominator by σ_v, Eq. 19.44 becomes

$$\text{var}(v) = \frac{\lambda\sigma_v^2}{8\sqrt{\pi}\,\sigma_v T_0} = \frac{\sigma_v^2}{4\sqrt{\pi}\,\sigma_f T_0} \tag{19.45}$$

Thus, it is seen that the variance of the mean velocity estimate $\hat{v}$ is directly proportional to the variance of the doppler spectrum and inversely proportional to the number of independent samples, $\sigma_f T_0$. Note also that $\text{var}(\hat{v})$ is proportional to λ, indicating that, for the same processing time T_0 and for the same σ_v, the variance of the estimate can be reduced by reducing the wavelength, which increases the number of independent samples.

Equations 19.42–19.45 are applicable in high signal-to-noise ratio cases. Uncertainties in the spectral moment estimates result from the limited observation time of the narrow band random process characterizing the meteorological echo. Any noise in the measurement further increases the uncertainty. Zrnić[85] gives the more general following expression for the variance of the mean frequency estimate f for the pulse-pair estimation technique and a gaussian-shaped spectrum

$$\text{var}(\hat{f}) = \frac{1}{8\pi^2 T_0 \rho^2(T) T}\left\{2\pi^{3/2}\sigma_f T + \frac{N^2}{S^2} + 2\frac{N}{S}\left[1-\rho(2T)\right]\right\} \tag{19.46}$$

where ρ is the correlation coefficient and N/S is the noise-to-signal ratio. Equation 19.46 applies a single PRF with interpulse period T and assumes that all pulses in the interval T_0 are used in the estimation algorithm. It reduces exactly to Eq. 19.43 for a large SNR and for narrow spectra, i.e., $\rho(T) \approx 1$. The reader is referred to Zrnić[85] for further details regarding the estimation of other moments of the doppler spectrum.

Pulse Compression. Pulse compression has been infrequently used for meteorological applications because short pulse peak power has not been a limitation on weather radar system performance. Keeler and Frush,[92] however, described how atmospheric distributed targets may be treated as frozen (fixed) "slabs" of scattering centers such that they approximate layers of nonfluctuating scatterers as the coded radar pulse propagates through them. Thus, each scattering slab produces a return signal that can be compressed by the compression filter in the same way as for individual point targets. They pointed out that pulse compression can benefit rapid-scanning radar applications in which the desired dwell time of each scattering volume is much shorter than the decorrelation time of the weather echoes. In these cases, averaging pulse compressed range measurements, instead of integrating over a longer dwell time, provides the large number of independent samples needed for accurate measurement in a very short beam observation time. Similarly, the effective range resolution over which averaging is performed can be flexible to meet changing observational requirements—short range intervals for highly structured heavy rainfall storms (typically convective) and longer range intervals for more uniform but weaker rainfall (typically stratiform). In other situations where signals are very weak (such as for wind profiler applications and clear air boundary layer observations), pulse compression may be used to increase system sensitivity[93] by increasing the average power of the system using long pulses while processing for any desired range resolution. Range resolution and SNR are independently determined by the pulse compression waveform design parameters.

A note of caution is in order when considering pulse compression for meteorological radars, relating to the matter of range sidelobes. Careful design is necessary to minimize these sidelobes, just as antenna sidelobes should be minimized, in order

to mitigate the effects of interpretive errors caused by wide dynamic range distributed weather targets.[94,95] Amplitude, frequency, and phase shaping of the transmitted pulse along with special sidelobe suppression/compression filters in the receiver processor allow range sidelobes to be suppressed by more than 50 dB over a significant doppler interval—tens of meters per second.[96] These compression filters are not matched to the transmit waveform; therefore, the SNR degrades somewhat and possible loss of range resolution in the main beam may occur.[95] However, these losses are tolerable for many meteorological measurements of interest.

Whitening. Pulse compression involves increased transmit bandwidth that is frequently difficult to have allocated and may limit its use at S band and C band for operational applications. A fixed bandwidth technique for increasing range resolution and the number of independent samples is the process of sampling the transmit pulse at an interval several times shorter than its duration and using a linear predictive *whitening filter*[97] to remove the correlated information from the oversampled data.[98,99] This whitening process increases the number of independent samples for range averaging that allows improved parameter estimation accuracy but at a cost of significantly reduced SNR. The higher noise bandwidth in the receiver required by the higher rate sampling and the whitening filter noise enhancement cause the SNR to be reduced by a factor of approximately L^2 where L is the increased sampling factor.[100] Stated another way, the whitening filter effectively shortens the transmit pulse (by L) while passing the increased receiver noise power (by L) thereby trading range resolution for SNR. Fortunately, for precipitation echoes and typical weather radars, the SNR is quite high, and even after the whitening process, it remains high enough to effectively utilize the larger number of independent samples for improved base data estimates or improved range resolution measurements.[101] On the other hand, if the signal is weak, noise enhancement will dominate the whitening process and any advantage is lost. Highly important is the fact that the transmit pulse and its bandwidth are not changed; consequently, increased frequency allocation to make use of the higher range resolution is not an issue.

Short Dwell Time Spectra (Maximum Entropy). Spectral processing of weather echoes adds another degree of freedom (the frequency dimension) to the ability to discriminate signal from ground clutter, other artifacts, and noise, and to estimate meteorological parameters of interest. Most frequency domain processing requires relatively long IQ data sample sets for discrete Fourier transform analyses, window functions, and possible spectral averaging to obtain spectra suitable for quantitative processing.[102] Faster scanning radars used for sampling rapidly evolving storms requires spectrum analysis techniques that use short dwell time data sets of many fewer IQ data samples. Modern spectrum analysis techniques such as the Burg maximum entropy[103] and the Capon[104] maximum likelihood estimators allow using short dwell sampled data to obtain stable spectrum estimators. These techniques belong to the general class of autoregressive (AR) estimators whereby the observed data are modeled as all-pole filtered white noise rather than a weighted sum of sinusoids according to the Fourier model.[105] Multiple signals and ground clutter may be resolved the same as using Fourier estimators. These short dwell time AR spectra may then be used to estimate weather spectral moments the same as the Fourier data model spectrum estimators from which the weather parameters are derived.[87]

Processor Implementations. Modern meteorological radars use digital signal processing techniques on programmable platforms and interactive color displays

for quantitative precision in interpreting the weather echoes. Modern weather radars require large dynamic range to sense strong echo at short range and weak echo at long range. Thus, the receiver and processor designs attempt to maintain amplitude and phase linearity throughout that range by using a dynamic automatic gain control (AGC), whereby the receiver gain and phase are adjusted along the range interval through the use of rapidly switched attenuators or, more often, digital compensation. Clearly, such rapid switching in the receiver requires careful design in order to avoid the affects of switching transients. An approach that avoids transient effects is to use two parallel IF receiver channels, each with moderate dynamic range and fixed gains, and to sample the signal in the channel that is best matched to the signal strength.

In all those approaches, it is possible to achieve linear dynamic range of greater than 90 dB and to use floating-point digital processing. The reflectivity, mean doppler velocity, and spectrum width can all be estimated digitally from the floating-point samples. The processing can be performed with a dedicated digital signal processing computer or by fast general-purpose computers combined with special signal processing components using Digital Signal Processing (DSP) chips or field programmable gate array (FPGA) devices. Continued technology advances will permit advanced, but mature, signal processing algorithms such as modern spectral processing and adaptive filtering to be implemented.[87]

19.5 OPERATIONAL APPLICATIONS

As has been demonstrated, meteorological radars measure backscattered power and radial velocity parameters. The challenge to the radar meteorologist is to translate these measurements, their spatial distributions, and their temporal evolution into quantitative assessments of the weather. Serafin and Wilson,[1] among others, show how modern meteorological radars are used for forecasting the weather. The level of sophistication used in interpretation varies broadly, ranging from human interpretation of rudimentary displays to automatic algorithms and modern multidimensional displays to assist human interpretation. Expert system approaches[106,107] that attempt to reproduce human interpretive logical processes have been employed effectively. The degree to which automation can be applied is evident in the Nexrad radar system design that produces the automated meteorological products shown in Table 19.7.

TABLE 19.7 Partial List of Nexrad Automated Products

Base radar reflectivity (all ranges, each scan)
Base radial velocity
Base spectrum width
Composite (maximum at all heights) reflectivity
Precipitation echo tops
Severe weather probability
Velocity Azimuth Display (VAD) Wind Profile
Storm relative mean radial velocity
Vertical integrated liquid
Storm tracking information
Hail potential
Mesocyclone and Tornado Vortex Signature
Surface rainfall accumulation (1 hr, 3 hr, total storm)
Radar Echo Classifier

Precipitation Measurement. Among the more important parameters to be measured is rainfall, having significance to a number of water resource management problems related to agriculture, fresh-water supplies, storm drainage, and warning of potential flooding. The rainfall rate can be empirically related to the reflectivity factor by an expression of the form[108]

$$Z = aR^b \tag{19.47}$$

where a and b are constant and R is the rainfall rate, usually in millimeters per hour. Battan[20] devotes three full pages of his book to the listing of dozens of Z-R relationships derived by investigators at various locations throughout the world, for various weather conditions, and in all seasons of the year. The fact that no universal expression can be applied to all weather situations is not surprising when one notes that rainfall drop-size distributions are highly variable. For many conditions, the drop-size distribution can be represented by an exponential function

$$N(D) = N_0 e^{-\Lambda D} \tag{19.48}$$

where N_0 and Λ are constants and $\Lambda D_0 = 3.67$ where D_0 is the median drop diameter. If $N(D)$ is known, the reflectivity factor can be calculated from Eq. 19.8. By using the terminal-fall speed data of Gunn and Kinzer,[109] the rainfall rate can also be obtained and Z related to R as shown in Eq. 19.47. The exponential form of the drop-size distribution generally fits the average of several drop-size distributions that occur over different phases of convective and tropical rainfall and are averaged over space or time. However, the "gamma" drop-size distribution represents a better fit to instantaneous, natural variations of the drop-size distribution under different conditions. The gamma drop-size distribution[24] is given as

$$N(D) = N_0 D^{\mu} e^{-\Lambda D} \tag{19.49}$$

with $\mu > -1$ and $\Lambda D_0 = 3.67 + \mu$. The μ parameter controls the shape of the distribution and when $\mu = 0$, the distribution is exponential.

Clearly, a single-wavelength, single-polarization radar can measure only a single parameter Z and must assume Rayleigh scattering. Since the rainfall rate depends upon two parameters, N_0 and Λ, it is not surprising that Eq. 19.47 is not universal. Despite this fact, Battan[20] lists four expressions as being "fairly typical" for the following four types of rain:

$$\text{Stratiform rain}[110]\ Z = 200\ R^{1.60} \tag{19.50}$$

$$\text{Orographic rain}[111]\ Z = 31\ R^{1.71} \tag{19.51}$$

$$\text{Thunderstorm rain}[112]\ Z = 486\ R^{1.37} \tag{19.52}$$

$$\text{Snow}[113]\ Z = 2000\ R^{2.00} \tag{19.53}$$

Stratiform[53] refers to widespread, relatively uniform rain and uses the well-known Marshall-Palmer drop-size distribution. *Orographic*[53] rain is precipitation that is induced or influenced by hills or mountains, whereas *Thunderstorm*[53] rain is typical of convective precipitation systems. In each of the above expressions, Z is in mm^6/m^3 and R is in mm/h. In Eq. 19.53, R is the precipitation rate of the water-equivalent melted snow. Battan[20] gives a more complete treatment of this important topic.

Wilson and Brandes[114] give a comprehensive treatment of how radar and rain-gauge data can be used to complement one another in measurements of precipitation over large areas. They state that the radar determined storm cumulative precipitation measurements are expected to be accurate within a factor of 2 for 75% of the time. Accuracies over large areas can be improved to about 30% with the addition of a surface rain-gauge network. Although radar measures reflectivity aloft, the primary concern is rainfall estimation at the surface. Rain gauge measurements are frequently used to adjust the radar reflectivity values.[115] Zawadzki[116] describes many factors that influence rainfall measurement by radar. Joss and Lee[117]and others[118,119,120] use the Vertical Profile of Reflectivity (VPR) to estimate the surface rainfall rate with some success. Bridges and Feldman[121] discuss how two independent measurements (reflectivity factor and attenuation) can be used to obtain both parameters of the drop-size distribution and more precisely determine the rainfall rate.

Polarimetric Estimates. Dual polarization radar has been proven to yield improved rainfall estimates as well as other advantages for weather radar.[122] Polarimetric radar measurements take advantage of the flattened raindrop shape approximated by an oblate spheroid (a flattened sphere having a horizontal dimension greater than its vertical dimension), and its propagation and scattering characteristics, to obtain accurate rain-rate estimators. Seliga and Bringi[123] and Sachidananda and Zrnic[124] show how the co-polar measurements of Z at horizontal (Z_h) and vertical (Z_v) polarizations can produce two independent measurements and, therefore, provide more accurate rainfall-rate measurements than those related to simple reflectivity. Bringi and Chandra[24] give the following rain-rate estimators for measurements at 10 cm (S band). Other wavelengths have different constants because of Mie factors entering into the calculations and the wavelength dependence of phase measurements.

$$R(Z_h, Z_{dr}) = 0.0067\, Z_h^{\,0.93}\, Z_{dr}^{\,-3.43} \tag{19.54}$$

$$R(K_{dp}) = 50.7\, K_{dp}^{\,0.85} \tag{19.55}$$

$$R(K_{dp}, Z_{dr}) = 90.8\, K_{dp}^{\,0.93}\, Z_{dr}^{\,-1.69} \tag{19.56}$$

The reflectivity-based rain-rate estimators corresponding to the Marshall-Palmer ($Z = 200R^{1.6}$) and the common Nexrad WSR-88D ($Z = 300R^{1.4}$) relations are

$$R_{MP}(Z) = 0.0365\, Z^{0.625} \quad \text{for Marshall-Palmer} \tag{19.57}$$

$$R_{88D}(Z) = 0.0170\, Z^{0.714} \quad \text{for Nexrad WSR-88D} \tag{19.58}$$

The polarimetric precipitation measurements bring unique characteristics that address not only improved precipitation measurements, but also improved data quality characteristics.[125] The differential reflectivity between the horizontal and vertical polarizations (Z_{dr}) allows estimating the effective drop size, whereas differential phase (K_{dp}) gives additional, somewhat independent information on estimation of rainfall rates. K_{dp} is an especially important parameter when the radar is confronted with typical contaminants as beam blockage, ground clutter domination, and calibration errors. Furthermore, Ryzhkov and Zrnic[126] have shown that K_{dp}-based rainfall measurements are less dependent on the unknown and variable drop-size distribution (DSD) than power-based measurements. Furthermore, other inferences from the

polarimetric measurements allow one to estimate the parameters of the gamma function form of the DSD,[127] which may further improve rainfall estimation. And, self consistency constraints of the polarimetric precipitation covariance matrix impose relatively strict bounds on the absolute reflectivity measurements, allowing one to calibrate the reflectivity measurement and make improved rainfall rate estimates.[128] For short wavelength radars, one must consider Mie scattering effects when measuring rainfall rate by reflectivity measurement. Lhermitte[25] and Kollias et al.[129] give an analysis of Mie scattering and show that if the scattering is properly accounted, the rainfall measurements are accurate.

Polarimetric radars may be configured in different ways for different measurements. Research radars explore the depolarization ratios by alternating horizontally and vertically polarized pulses. However, the preferred configuration for operational radars is to transmit both horizontal and vertical polarizations simultaneously while independently receiving both orthogonal signals for polarimetric processing, allowing reception of the important co-polar quantities. The relative phase of each transmitted polarization is arbitrary, but the differential magnitude and phase of the co-polar received signals in each polarization are sufficient for the polarimetric measurements.

It is our opinion that no single topic in radar meteorology has received more attention than rainfall rate measurement. Although useful empirical expressions have evolved and dual-polarization techniques show significant improvements in accuracy, a completely satisfactory approach remains to be developed.

Polarimetric measurements by the WSR-88D radars will allow more accurate precipitation measurements and ice-water phase knowledge that is important for making improved forecasts, separating winter precipitation into regions of rain and snow, and detecting hazardous aircraft icing conditions.[130] Implementation of phase-coding algorithms allow these radars to separate multiple time around (or multiple trip) echoes[67,131,132] and to compute dealiased velocities in the range dealiased echoes[133]; to classify and separate precipitation echoes from artifacts as ground clutter, sea clutter, and insects echoes[72]; to accommodate flexible scan strategies; and to incorporate improved data resolution in range and azimuth. These performance enhancements are readily achieved by installing greater and more flexible signal and data processing power.

Severe Storm Warning. One of the primary purposes of weather radars is to provide timely warnings of severe weather phenomena such as tornadoes, damaging winds, flash floods, and landfalling hurricanes. Accurate long-term forecasting of the precise location and level of severity of these phenomena, through data assimilation and numerical weather prediction techniques, is beyond the present state of the art. Operational radars, however, can detect these phenomena and provide local warnings of approaching severe events; they can also detect the rotating mesocyclones in severe storms that are precursors to the development of tornadoes at the Earth's surface.[134] Ground-based coastal and airborne radars can also measure the severity of approaching hurricanes and define their most intense landfall positions for evacuation warnings.[135,136]

Tornado Detection. A single doppler radar can measure only the radial component of the vector wind field. Hence, exact measurements of vector winds at a point are generally not possible. However, rotating winds or vortices can be detected and

their intensities measured by simply measuring the change in radial velocity with azimuth angle, as shown in Figure 19.6. The radar scans in azimuth and detects a couplet in radial velocity at constant range. The azimuth shear is given simply by the expression

$$\frac{dv_r}{dx} \approx \frac{2v_r}{r\alpha} \tag{19.59}$$

where x is in the direction orthogonal to the radius of the circulation and α is the angle subtended by the circulation at range r.

Because mesocyclones may spawn tornadoes and can be many kilometers in diameter, radars with about 1° beamwidth have the spatial resolution to detect mesocyclones at ranges in excess of 60 km. It should be clear that any mean translational motion would change the absolute values of the measured radial velocities but would not affect the shear measurement. Azimuthal shear values of the order of 10^{-2} s^{-1} or greater and with vertical extent greater than the diameter of the mesocyclone are deemed necessary for a tornado to occur.[137]

Detection of the tornado vortex itself is not generally possible beyond about 30 km using typical weather radar beamwidths, since its horizontal extent may be only a few hundred meters. Detection of the radial shear, therefore, is not possible unless the tornado is close enough to the radar to be resolved by the beamwidth. However, in cases where the tornado falls entirely within the beam, the doppler spectrum width may be used to estimate tornadic intensity. In some cases, both a mesocyclone and its incipient tornado can be detected. Wilson and Roesli[138] show an excellent early example of a tornado vortex signature (TVS) embedded within a larger mesocyclone.

Microbursts. Fujita and Caracena[139] first identified the microburst phenomenon as the cause of an airliner crash that took place in 1975. The microburst and its effects on an aircraft during takeoff or landing are depicted in Figure 19.7. The microburst is simply a small-scale, short-duration downdraft emanating from a convective storm. This "burst" of air spreads out radially as it strikes the ground, forming a ring of *diverging*[53]air about 0.3 to 1 km deep and of the order of 2 to 4 km in diameter with divergent winds greater than 10 m/s and lasting less than 20 minutes. Aircraft penetrating a microburst first experience an increase in head wind and then a continuous, performance-robbing decrease in head wind, which can cause the plane to crash if encountered shortly before touchdown or just as the aircraft is taking off. More complete descriptions of microbursts and their effects on aviation safety are given by Fujita[140,141] and McCarthy and Serafin.[142]

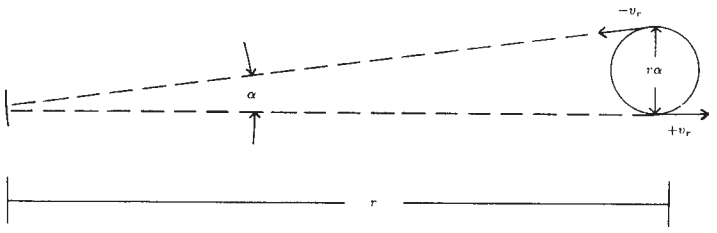


FIGURE 19.6 Measurement of rotation or azimuthal wind shear in a mesocyclone, a rotating wind parcel. The azimuthal shear is given by $\Delta v/\Delta x = 2v_r/r\alpha$. (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

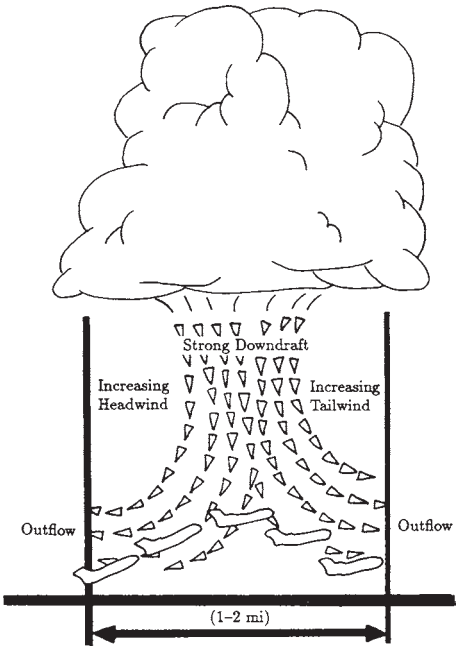


FIGURE 19.7 Depiction of a microburst and its effect on an aircraft during takeoff. The loss of airspeed in the divergent wind field near to the ground is extremely hazardous. A similar loss of airspeed near the ground occurs during landing. (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

Microburst detection, like tornado detection, is accomplished by estimating shear. However, in the case of the microburst, it is the radial shear of the radial velocity that is typically measured. Human interpretation of microburst signatures in color-enhanced radial velocity displays is easily accomplished with trained observers[50] and automatic detection has been implemented on the TDWR radar system. Radial velocity differences of 10 to 50 m/s are observed in microbursts. A radial velocity difference of more than 25 m/s over the length of a jet runway (≈3 km) is a serious concern.

One principal problem concerning microbursts is their short lifetimes, which are on the order of 15–20 minutes but the duration of peak intensity is only 1 or 2 minutes. Field research[143] has clearly demonstrated that a few minutes advance warning using doppler radar can be achieved. Operational microburst detection radars use this automated detection algorithm with high performance ground clutter mitigation techniques since the phenomenon occurs near the ground and oftentimes in very light or no precipitation.

C band is the preferred operational frequency for several reasons. First, a C-band antenna will have a smaller beamwidth than an S-band antenna of the same size and allow improved air flow measurements with strong clutter suppression near the airport surface. Second, since short range detection is important, radar attenuation effects are not a primary concern. Third, C band offers improved signal-to-clutter

performance since the large clutter targets are limited in radar cross section by Mie scattering, whereas the atmosphere-borne wind tracing scatterers are small and obey Rayleigh scattering cross section physics. X-band radars have been considered to some degree but the more severe attenuation that can occur in very heavy rain will limit performance unless the radar is located very near to the airport runways. In the mid-1990s the C-band Terminal Doppler Weather Radar (TDWR) system was installed at 45 major airports to detect and warn aircraft of hazardous wind shear conditions, approaching gust fronts that may affect the airport approach and departure configuration, and microbursts.[7] This network of radars along with much improved pilot training and awareness has all but eliminated aircraft accidents caused by microbursts and strong wind shear.

Hail. The Nexrad radar makes use of a hail-detection algorithm that combines a high reflectivity factor with echo height and upper-level divergent radial velocity winds to detect the occurrence of hail. Polarimetric radar techniques improve quantitative hail detection. Bringi et al.,[15] Aydin et al.,[144] and Illingworth et al.[145] proposed a hail detection technique using differential reflectivity measurements. This technique depends upon the fact that for tumbling hail the differential reflectivity, the ratio of horizontal to vertical reflectivity, is near unity (≈0 dB). This differs sharply from heavy rain, where this ratio can be as large as 5 dB because large water droplets are horizontally oriented. The combination of absolute reflectivity factor and the dual polarization differential reflectivity gives unique signatures for hail and heavy rain, each of which is characterized by a high reflectivity factor. The difference in the differential reflectivity signatures is easily explained. Large raindrops assume flatter shapes (approximated by oblate spheroids) as they fall and thus scatter back the horizontally polarized electric field more strongly than the vertically polarized electric field. Hailstones, having an irregular shape, physically tumble while they fall and, therefore, exhibit no preferred orientation on average. Thus, the horizontal and vertical scattered fields have nearly the same average value.

Wind Measurement. Lhermitte and Atlas[146] were the first to show how a single doppler radar can be used to measure vertical profiles of the horizontal wind field when precipitation is present. This technique is most accurate if the wind field is uniform in the region scanned by the radar. The method depends upon an analysis of the radial velocity measured during a complete scan in azimuth with a single fixed elevation angle. At any slant range r, the height of the measurement is $r \sin \alpha$ and the radius of the region scanned is $r \cos \alpha$, where α is the elevation angle as depicted in Figure 19.8. If β is the azimuth angle, V_h is the horizontal wind speed, and V_f is the fall speed of the particles, then the radial velocity at range r is given by

$$V_r(\beta) = V_h \cos \beta \cos \alpha + V_f \sin \alpha \tag{19.60}$$

A harmonic analysis (least square fitting the amplitude, phase, and offset of a sinusoid) can be used to obtain V_h, the horizontal wind speed, the wind direction where cos β is maximum, and V_f, the mean particle fall speed all plotted as a function of height. This technique is referred to as the Velocity-Azimuth-Display (VAD). Browning and Wexler[147] later showed how the VAD technique could be extended to measure other parameters of the wind field including wind-field divergence and wind-field deformation by employing an extended harmonic or Fourier series analysis. The VAD has been implemented on the Nexrad as a standard product that can be used in precipitation and

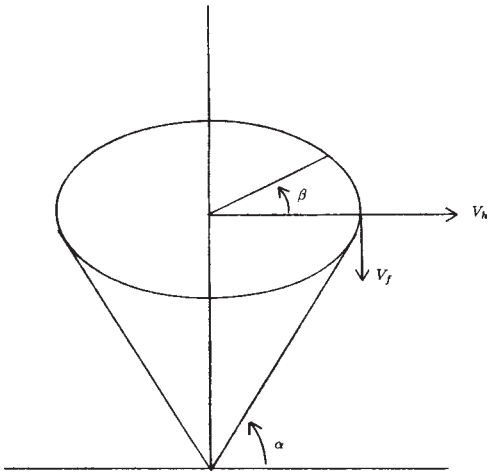


FIGURE 19.8 Velocity-Azimuth-Display geometry for measuring horizontal wind with a single doppler radar. Measurement of the radial velocity for a complete azimuthal scan (β) at elevation angle α permits an estimate of the vertical profile of horizontal winds. (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

frequently in clear air. The VAD technique is most often applied to wind profiler radars that point vertically and step scan at relatively large elevation angles. An alternative determination of boundary layer wind fields using a single weather radar has found success using an echo tracking technique.[148]

Thunderstorm Prediction. Wilson and Schreiber[149] illustrate how meteorological doppler radar can be used to detect locations where new thunderstorm development is likely to occur. Many weather radars have sufficient sensitivity to detect discontinuities of clear air echoes in the lower 2 to 4 km of the atmosphere out to 50 or 100 km. This detection occurs principally in the summer months when the backscattering mechanism is caused by insects in the lower levels of the atmospheric boundary layer and sometimes may also be due to Bragg-scattered refractive index inhomogeneities. Wilson and Schreiber have found that about 90% of the thunderstorms that occur in the Front Range of the Rockies in the summertime develop over such "boundaries" between two different air masses. Since these boundaries can be detected before any clouds are present and it is possible to infer the air mass convergence (or coming together of two air masses) where insects are forced to accumulate along these boundaries by doppler radar measurements, more precise prediction of thunderstorm occurrence is possible. From the radar designer's standpoint, such applications dictate the use of antennas with very low sidelobes (along with low-phase noise transmitters and receiver components) and signal processors having significant ground clutter rejection capability. The Nexrad radar system, with its high quality antenna and 50 dB of clutter rejection, is well suited to this important operational task.

Refractivity and Water Vapor Measurements. Conventional weather radar processing is generally designed to emphasize precipitation and wind-field measurements

and to suppress ground clutter. However, it is known that radar beam bending caused by atmospheric refraction and the resulting anomalously propagated ground clutter echo gives an indication of the vertical profiles of temperature and moisture in the intervening lower atmosphere. Moreover, measuring the propagation speed of the radar pulse between normally propagated clutter targets (those directly viewed by the radar) gives an estimate of the refractive index of air along this propagation path. By measuring the absolute phase of the radar signal received from stationary ground clutter targets and comparing them with reference absolute phase measurements under known refractive conditions, one can measure the near surface propagation speeds of the radar pulse along these paths. Then, the refractive index, or refractivity, of the atmosphere of these atmospheric paths may be determined.[17] Refractivity is a function of temperature, pressure, and moisture content. Consequently, if surface temperature and pressure are independently available, as they frequently are, then the refractive index measurements may be converted to spatial fields of water vapor in the surface boundary layer. Such water vapor measurements are critical for obtaining accurate forecasts using numerical models and simulations of the atmosphere.[150] This experimental measurement technique can be used to ranges of about 40 km where the Earth's curvature prevents routine observation of ground clutter targets and is being considered for operational use in Nexrad.

19.6 RESEARCH APPLICATIONS

Operational meteorological radars are designed for reliability and simplicity of operation while providing the performance needed for operational applications. Research radars are considerably more complex, since cutting-edge research requires more detailed and more sensitive measurements of a multiplicity of variables simultaneously. In the research community, multiple-parameter (polarization and wavelength) radar studies, multiple doppler radar network studies, and a new generation of airborne and spaceborne radars are receiving considerable attention.

Dual Polarization/Wavelength Radar. It is clear that polarimetric doppler radars provide a significant increase in the useful information that can be obtained from meteorological targets. The detection of hail and more accurate rainfall estimates has primary significance. Dual polarization measurements at multiple wavelengths provide even more information related to the eventual interpretation of the size distributions, water-phase states, and hydrometeor (water or ice particles) types in different classes of precipitation and clouds. The capabilities of multiple-parameter meteorological radars are presented in Bringi and Hendry,[14] Bringi and Chandra,[24] and in the collection of papers edited by Hall.[151] Whereas longer-wavelength radars are necessary for the study of severe storms, shorter wavelength millimeter-wave radars are useful for sensing and probing newly developing clouds. Researchers often need a wide range of these capabilities simultaneously. From the radar engineering standpoint, the challenge is considerable, requiring radar designers to develop coherent waveform, polarization-diverse, and wavelength-diverse radars. As noted, there exist several polarimetric research radars and operational radars in the world.[26]

Multiple Radars. A single doppler radar measures only a single radial component of velocity. Lhermitte[152] was among the first to describe how two or more doppler radars could be used, scanning together, to obtain the full three-dimensional air motion

fields in precipitation. This pioneering work led the way toward the use of networks of doppler radars for studies of individual clouds to examine the three-dimensional structure of vector air motion in precipitation. A point should be made here regarding the use of two doppler radars for measurements of three-dimensional winds. Since, in principle, two independent looks can measure only two components of vector air motion, the assumption of atmospheric mass continuity is invoked. This *equation of mass continuity*[53] ($\nabla \cdot \bar{V} = 0$) is used to obtain the third-dimensional component, where $\bar{V}$ is the vector air motion and is constrained to be zero at the surface. The vertical air motion is calculated from vertical integration of the mass continuity equation.

Figure 19.9 illustrates an air motion field obtained by two doppler radar observations in an individual convective storm cell. Shown are the horizontal vector fields in a plane approximately 100 m above the surface. The phenomenon being measured is a low-level divergent outflow (or microburst) just to right of the center. Figure 19.10 shows another example of precipitation intensity and air motion fields superposed on a photograph of a convective thunderstorm cell. The data are taken from three doppler radars spaced about 40 km apart.

A developing network of short range doppler radars not range limited by the Earth's curvature will provide more detailed observations close to the surface than the relatively widely spaced WSR-88D network. The Collaborative, Adaptive Sensing of the Atmosphere (CASA) radar network will include many inexpensive, low power, scanning, short wavelength radars mounted on towers of opportunity, primarily cell phone towers that blanket much of the U.S.[153,154] Existing equipment on these towers will

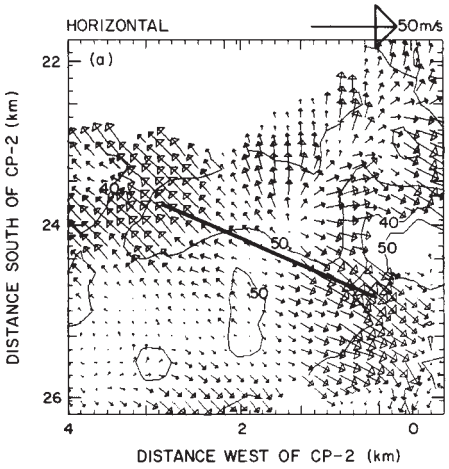


FIGURE 19.9 Vector wind fields in a horizontal plane derived from dual-doppler radar observations of a summertime convective storm near Denver, Colorado. The dark, solid line is shown to indicate the length of a typical jet aircraft runway. The divergent outflow from a microburst shows a strong 15 m/s headwind changing to a similarly strong tailwind over the length of the runway. CP-2 is the former research radar operated by NCAR. (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

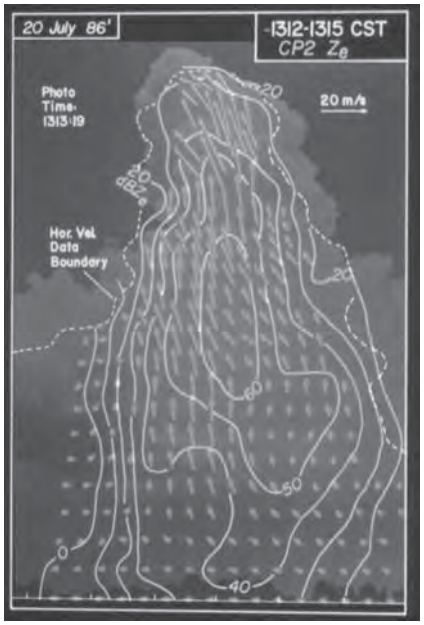


FIGURE 19.10 Radar reflectivity contours and air flow vectors superposed on photograph of a rapidly growing Alabama thunderstorm cell. The data depicts a 30 m/s updraft originating in an extremely strong 60 dBZ rain or hail core. The wind vectors are taken from combining data from three nearby doppler radars. (*Courtesy of American Meteorological Society. From D.E. Kingsmill and R. M. Wakimoto, "Kinematic, Dynamic, and Thermodynamic Analysis of a Weakly Sheared Severe Thunderstorm over Northern Alabama,"* Monthly Weather Review, *vol. 119, p. 272, 1991*)

provide a shared infrastructure for the CASA radars. This radar system will include intelligent decision making for scanning and tracking important low-level atmospheric features that are deemed critical for the aviation, weather forecasting, transportation, and local emergency response users. The CASA radars can communicate with each other for collaborative use of the data and adaptively change their operating parameters to best meet the prescribed needs at a given time. Depending on the final costs of deployment and operation, CASA networks may be developed only around critical areas such as urban areas, airports, etc.

Rapid Scan (Phased Array) Radar. Doppler weather radars that use phased array antennas and complex waveform design may be applied to some of the more difficult radar meteorology observations. The use of multiple doppler radars has provided dramatic new information on the internal winds in large precipitating systems—information that can be obtained in no other way. Despite the power of this technique, the spatial resolution in the derived three-dimensional motion fields is generally not better than

of the order of 1 km. There are several reasons for this. The finite beamwidth limits the resolution available at longer ranges. At shorter ranges, the large solid angle that must be scanned in order to cover all regions of a storm requires total scanning times of the order of 3 to 5 min even for well-situated storms. This is a consequence of the on-target dwell time necessary for accurate measurements. Finally, the storm itself is evolving and moving during this measurement time, thus complicating the data registration in space and time.

Some operational and many research applications require faster scanning than conventional mechanically scanned radars can provide.[155] These applications include longer lead times for tornado warnings, the study of finer scale storm features, interactions between the internal motions and hydrometeor growth processes in the storms, and studies of electric charge separation in clouds. Brook and Krehbiel[156] were among the first to discuss a very rapid-scanning radar (although nondoppler) for effectively obtaining snapshots of convective storms. Keeler and Frush[157] discuss design considerations for a more capable rapid-scanning doppler radar.

Any rapid-scanning approach generally must encompass two features: (1) the transmit waveform must have a relatively large bandwidth to increase the number of independent samples available in the desired spatial resolution cell; and (2) the antenna must be rapidly scanned throughout the desired atmospheric volume either mechanically or electronically. Mechanically scanned single frequency (standard) weather radars use a long dwell time to acquire an adequate number of independent samples to accurately estimate the spectral moment data used for meteorological measurements. A large bandwidth system (e.g., pulse compression radar) can acquire independent samples in range in a short dwell time that can be averaged, thus reducing the dwell time for each beam and reducing the total volume coverage scan time. Fast mechanical scanning (>60 deg/sec) produces deleterious spectrum broadening effects[157] that electronic step scanning easily avoids since the beam is fixed in space during the dwell time. A favored approach is to utilize a one-dimensional phased array that is electronically scanned in elevation while slowly rotating in azimuth.[158] In this way, a full hemispheric volume can be covered in 1–2 minutes and smaller sector volumes may be covered in less than 1 minute.[159] Several military radars and aviation radars utilize electronically scanned beams but the expense of a fully capable system has prevented more than only a few meteorological radar systems from being designed and built.[160,161] Rapid-scanning phased array radars have also been studied extensively in Europe.[162]

An alternative approach is to employ digital beamforming or frequency-steering techniques to transmit and simultaneously receive multiple beams using parallel receivers. Such military radars have been operated for several decades[163] but none of these systems were ever designed or used for weather measurements. A mechanically scanned multiple beam weather radar system using simultaneous frequency scanned beams has been designed and constructed on a small truck. The Rapid Doppler on Wheels (Rapid-DOW) X-band radar, shown in Figure 19.11, uses six simultaneous frequency-steered beams separated by a few degrees emanating from a slotted waveguide phased array and independent receivers to acquire a volumetric coverage of convective storms in 1–3 minutes.[164] Pulse compression is not used. The rapid data acquisition speed of this system has the potential to give new insights into thunderstorm evolution, hurricanes at landfall, hail formation plus microburst, gust front and tornado genesis mechanisms that may likely evolve into an operational capability.

In the early 1990s, the FAA led an effort to develop dual-use weather and aircraft surveillance radar based on electronic scanning and pulse compression techniques,[165]

FIGURE 19.11 Rapid-DOW (Doppler on Wheels) is a mobile X-band radar that uses six simultaneous beams to cover a volume of atmosphere in a much shorter volume coverage time than a mechanically scanned single beam radar. This rapid update scan is important for measuring violent convective storms, especially tornadoes as shown in photo. (*Photo courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

and an effort beginning a decade later continued the development of phased array radars for civil applications.[166] Radars initially designed for military applications were modified for weather detection and successfully demonstrated the concept.[167] In the early 2000s, a joint effort between the U.S. Navy, the National Weather Service, the National Severe Storms Laboratory, and the University of Oklahoma led to the development of the so-called National Weather Radar Testbed facility in Norman, Oklahoma, by combining a single panel SPY-1 phased array radar on a rotating pedestal with a WSR-88D transmitter and custom receiver.[168] Although this particular radar is not likely to be the preferred design for a future weather radar, the Oklahoma Phased Array Radar testbed can be used to explore electronic scanning strategies together with new pulsing and processing concepts that may lead to future development of a phased array weather radar system.

Airborne Radars. Whereas commercial aviation weather radars are nose mounted X-band radars for severe weather and turbulence detection and avoidance, airborne research radars must have relatively more complex architectures[169] in order to make more sensitive, high-resolution measurements. This powerful technique permits the use of a mobile platform, which therefore allows measurements over regions not accessible by ground-based systems. Moreover, the mobility of the aircraft permits longer term observations of rapidly moving but long-lived storms and cloud systems, thereby allowing more complete studies of evolution during various phases of the system.[170]

NCAR's Eldora airborne doppler radar system,[171] shown in Figure 19.12, consists of two slotted waveguide fixed-beam antennas mounted in the tail of the P-3 aircraft, which is operated by the Naval Research Laboratory and covered by a rotating

FIGURE 19.12 Tail-mounted ELDORA doppler weather radar operated by the National Center for Atmospheric Research on the Naval Research Laboratory P-3 research aircraft (*Courtesy of University Corporation for Atmospheric Research © 2007, Boulder, CO*)

radome—a roto-dome. One beam points forward about 18° and the other aft by the same angle, thereby obtaining two radial components of common target volumes. With such a system, each antenna scans in a conical surface—one cone pointing forward, the other one rearward—thus permitting synthesis of a dual-doppler radar system along the aircraft track. The aircraft is flown alongside storms to synthesize dual-doppler observations and, therefore, to obtain vector winds. Because the aircraft need not fly orthogonal tracks, the time required for measurements of cloud systems is dramatically reduced as are the errors in these measurements. Moreover, severe storms (which could otherwise not be penetrated along an orthogonal track) and hurricanes (which can be penetrated by such aircraft) can be observed fully by an aircraft outside the regions of severe weather.

Various airborne radars exist in the research community for different applications and user groups. NASA has developed the dual wavelength (3 cm and 3 mm) EDOP radar for the high altitude ER-2 aircraft[172] and a W-band 94-GHz airborne cloud radar for the DC-8 research aircraft.[173] NOAA has operated two X-band radars similar to the ELDORA radar for hurricane tracking and research.[174] JPL developed the ARMAR radar[175] to test space radar concepts for the Tropical Rainfall Measurements Mission (TRMM). The University of Wyoming uses a 3-mm cloud radar mounted on a King Air aircraft.[176] In the mid-2000s, NCAR acquired the High-performance Instrumented Airborne Platform for Environmental Research (HAIPER, a Gulf Stream G-5 medium-size jet aircraft) and designed a dual wavelength (8 and 3 mm) removable pod-mounted radar for cloud studies.[177] The Canadian National Research Council has developed a similar airborne radar system.[178]

Spaceborne Radars. Among the more significant challenges facing researchers today is the need to make global measurements of precipitation. Understanding global climate requires that quantitative measurements of precipitation be made throughout the world, particularly in the tropics and over the oceans. Satellite observations appear

to offer the only practical mechanism for obtaining these measurements. The TRMM satellite was launched in 1997 carrying the K_u-band single frequency Precipitation Radar (PR)[179] and a 2.4 m single beam array antenna that is steered 17° on either side of the spacecraft track. Its relatively low inclination orbit at 350 km altitude provides tropical precipitation measurements with 250 m range resolution and a 4.5 km footprint over a 250 km swath. A TRMM follow-on program, the Global Precipitation Measurement (GPM) program, envisions extending precipitation coverage to the mid-latitude (65° *N* and *S* latitude) flying in a 250 km altitude orbit and using dual wavelength precipitation radar (DPR)[180] at K_u and K_a bands for more accurate rainfall estimates using attenuation techniques.[181,182] The two radars will have matched beams from two slotted waveguide array antennas and provide coverage under the spacecraft track similar to TRMM.

CloudSat is a satellite launched in 2006 flying a W band (3 mm) cloud profiling radar (CPR) orbiting the Earth in a sun synchronous orbit at an altitude of about 700 km.[183] The transmitter is an Extended Interaction Klystron (EIK) high-power amplifier generating a 3.3 μsec monochromatic pulse having peak power of 1.7 kW. The antenna is a 1.85 m diameter reflector offset fed by a quasi-optical transmission line and produces a 0.12° beam with extremely low sidelobes. These radar design parameters allow an exceptional sensitivity of –26 dBZ at the Earth surface. CloudSat orbits in formation with four other satellites as part of the so-called A-train constellation of satellites that provide combined radar, lidar, and radiometric measurements for Earth studies. The CPR on-board CloudSat has 500 m vertical resolution with a 1.4 km footprint and is similar to the NASA Airborne Cloud Radar that has flown for several years on board the NASA DC-8 aircraft[173]. Combining these high-resolution cloud measurements with high sensitivity is one of the technical goals for acquiring new information regarding cloud effect on the Earth's climate.

Clear-Air Wind Profiling Radars. Another form of doppler radar that has become widely used, especially in the research community, is the so-called wind profiling radar,[10] or "wind profiler." Wind profilers usually take the form of VHF and UHF multiple fixed-beam systems, pointing vertically and at angles approximately 15° from the zenith to infer profiles of the horizontal wind averaged over the area of measurement. Such radars can make doppler measurements throughout the range of altitudes from a few hundred meters to 20 km or more above the surface, depending upon the wavelength selected and the power-aperture product available. These radars have the ability to measure high resolution winds continuously, which permits the observation of smaller scale temporal and spatial wind-field features that cannot be obtained from the global 12-hour rawinsonde (balloon launch) network. These smaller scale measurements are important for understanding local and regional weather and for effective forecasting on these scales.

Very powerful radars of this type are referred to as Mesosphere, Stratosphere, Troposphere (MST) radars because of their ability to make measurements throughout most of these atmospheric regions up to 60–100 km in altitude. Several major MST radar facilities located at facilities around the world operate at VHF frequencies around 50 MHz and observe upper atmospheric (tropospheric and lower stratospheric) winds or the higher level stratospheric and mesospheric winds. Shorter wavelength UHF wind profilers operating at 400–450 MHz sense atmospheric winds up to 20–25 km and these "tropospheric wind profilers" are the most widely used for operational weather observations. UHF wind profilers operating at 915 MHz in the U.S. and 1200–1300 MHz

in Europe cover the lower atmospheric winds up to 3–5 km (or a few km higher with larger antennas) where the strong moisture fluctuations present in the atmospheric boundary layer provide strong scattering signatures at these shorter wavelengths. These UHF boundary-layer wind profilers are typically used for air pollution monitoring and warnings as well as various research applications.

These clear-air radars receive energy backscattered from refractive index inhomogeneities caused by naturally occurring atmospheric turbulence.[184] The antenna systems usually take the form of phased arrays that form beams several degrees wide that are switched to 3, 4, or 5 nearly vertical beams for 1–2 minutes each and measure vertical profiles of wind every 5–30 minutes. The antennas are frequently of the coaxial collinear (co-co) style for those radars and Yagi array elements for those at the higher frequencies. The higher frequency UHF profilers typically use either Yagi or microstrip patch array antennas. Transmitters are generally in the form of high-powered, coherent transmitting tubes or solid-state amplifiers. At the University of Kyoto, Japan, the antenna-transmitter system consists of 475 crossed Yagi radiating elements, each with its own solid-state transmitter.[185] This approach allows for very flexible electronic scanning of the beam. NOAA operates a network of over thirty 404 and 449 MHz wind profilers in the central United States using solid-state transmitters that supply continuous wind profiles up to 20 km for improved weather forecasts and current upper air wind information for aviation applications.[186]

It is important to recognize that three-beam doppler systems can accurately measure horizontal winds in all three velocity components if the wind is uniform. Four- and five-beam systems allow one to determine the quality of the measurements by detecting the presence of a nonuniform wind flow. Carbone, Strauch, and Heymsfield[187] and Strauch et al.[188] address the issue of wind measurement error in detail. The reader is referred to the review paper by Röttger and Larsen[10] for a thorough treatment of wind-profiler technology and to the sequence of Tropospheric Profiling Symposia for tropospheric profiling applications of winds and other meteorological parameters.[189]

Spaced Antenna Techniques. An additional advantage of these long wavelength radars is their ability to measure not only the radial vertical wind components directly, but also the mean horizontal wind that is transverse to the vertical beam without scanning that beam off zenith. These profilers use the so-called spaced antenna techniques with multiple receivers to process the amplitude and phase differences of the echo structures as they translate over two adjacent antennas (usually subarrays of the same array antenna) to measure components of the horizontal or transverse wind.[190] In this manner, two orthogonal subarrays can measure components of the horizontal wind using cross-spectral or correlation processing techniques.[191] Since the measurement is made in pairs of overlapping beams directly above the radar, it is no longer necessary to assume or require horizontal homogeneity of the wind field in the larger area above the radar and/or the long integration times necessary to assure this homogeneity. These spaced antenna techniques are most frequently used when high spatial and temporal resolution measurements are required such as estimating detailed boundary layer turbulence fields. Zhang and Doviak have investigated using the dual beam–spaced antenna technique implemented with an electronically scanned phased array radar to estimate the transverse, as well as the radial, wind components at arbitrary scan angles.[192]

REFERENCES

1. R. J. Serafin and J. W. Wilson, "Operational weather radar in the United States: Progress and opportunity," *Bull. Am. Meteorol. Soc.*, vol. 81, pp. 501–518, AMS, Boston, 2000.
2. R. J. Serafin, "New nowcasting opportunities using modern meteorological radar," in *Proc. Mesoscale Analysis Forecast. Symp.*, European Space Agency, Paris, 1987, pp. 35–41.
3. T. D. Crum and R. L. Alberty, "The WSR-88D and the WSR-88D Operational Support Facility [Now Radar Operations Center]," *Bull. Am. Meteorol. Soc*., vol. 74, pp. 1669–1687, 1993.
4. T. D. Crum, R. E. Saffle, and J. W. Wilson, "An update on the Nexrad program and future WSR-88D support to operations," *Weather and Forecasting*, vol. 13, pp. 253–262, 1998.
5. J. McCarthy, J. Wilson, and T. T. Fujita, "The Joint Airport Weather Studies (JAWS) project," *Bull. Am. Meteorol. Soc.*, vol. 63, pp. 15–22, 1982.
6. M. Michelson, W. W. Shrader, and J. G. Wieler, "Terminal doppler weather radar," *Microwave J.*, vol. 33, pp. 139–148, 1990.
7. J. G. Wieler and W. W. Schrader, "Terminal Doppler Weather Radar (TDWR) system characterizations and design constraints," in *25th Int. Conf. on Radar Meteorol.* AMS, 1991, pp. J7–J9.
8. National Research Council, *Assessment of Nexrad Coverage and Associated Weather Services*, Washington, DC: National Academy Press, 1995.
9. H. W. Baynton, R. J. Serafin, C. L. Frush, G. R. Gray, P. V. Hobbs, R. A. Houze, Jr., and J. D. Locatelli, "Real-time wind measurement in extratropical cyclones by means of doppler radar," *J. Appl. Meteorol*., vol. 16, pp. 1022–1028, 1977.
10. J. Röttger and M. F. Larsen, "UHF/VHF radar techniques for atmospheric research and wind profiler applications," Chapter 21 in *Radar in Meteorology*, Atlas (ed.) Boston: AMS, 1990, pp. 235–281.
11. V. Chandrasekar, R. Meneghini, and I. Zawadzki, "Global and local precipitation measurements by radar," Chapter 9 in *Radar in Atmospheric Science: A collection of essays in honor of David Atlas*, R. Wakimoto and R. Srivastava (eds.), Meteorological Monograph, vol. 30, Boston: AMS, 2003, pp. 215–236.
12. J. Wurman, J. Straka, E. Rasmussen, M. Randall, and A. Zahrai, "Design and deployment of a portable, pencil-beam, pulsed, 3-cm doppler radar," *J. Atmos. Oceanic Technol.*, vol. 14, pp. 1502–1512, 1997.
13. D. S. Zrnic, "Weather radar polarimetry: Trends toward operational applications," *Bull. Amer. Meteorol. Soc.*, vol. 77, pp. 1529–1534, 1996.
14. V. N. Bringi and A. Hendry, "Technology of polarization diversity radars for meteorology," Chap. 19 in *Radar in Meteorology*, Atlas (ed.), Boston: AMS, 1990, pp. 153–189.
15. V. N. Bringi, T. A Seliga, and K. Aydin, "Hail detection with a differential reflectivity radar," *Science*, vol. 225, pp. 1145–1147, 1986.
16. J. Vivekanandan, D. S. Zrnic, S. M. Ellis, R. Oye, A. V. Ryzhkov, and J. Straka, "Cloud microphysics retrieval using S-band dual polarization radar measurements," *Bull. AMS*, vol. 80, pp. 381–388, 1999.
17. F. Fabry, C. Frush, I. Zawadzki, and A. Kilambi, "On the extraction of near-surface index of refraction using radar phase measurements from ground targets," *J. Atmos. Ocean. Tech.*, vol. 14, pp. 978–987, 1997.
18. P. H. Hildebrand and R. K. Moore, "Meteorological radar observations from mobile platforms," Chapter 22 in *Radar in Meteorology*, Atlas (ed.), Boston: AMS, 1990, pp. 287–322.
19. R. J. Serafin and R. Strauch, "Meteorological radar signal processing in 'air quality meteorology and atmospheric ozone,'" *American Society for Testing and Materials*, pp. 159–182, Philadelphia, 1977.
20. L. J. Battan, *Radar Observation of the Atmosphere*, Chicago: University of Chicago Press, 1973.
21. D. Atlas (ed.), *Radar in Meteorology*, Boston: AMS, 1990.
22. B. R. Bean, E. J. Dutton, and B. D. Warner, "Weather effects on radar," in *Radar Handbook*, 1st Ed., M. Skolnik (ed.), New York: McGraw-Hill Book Company, 1970, pp. 24-1–24-40.

23. R. J. Doviak and D. S. Zrnić, *Doppler Radar and Weather Observations*, 2nd Ed., Mineola, NY: Dover Publications, 2006.
24. V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler Weather Radar: Principles and Applications*, Cambridge, UK: Cambridge Univ. Press, 2001.
25. R. M. Lhermitte, *Centimeter & Millimeter Wavelength Radars in Meteorology*, Miami: Lhermitte Publications, 2002.
26. R. E. Rinehart, *Radar for Meteorologists*, 4th Ed., Columbia, MO: Rinehart Publications, 2004.
27. "Special issue on radar meteorology," *IEEE Trans. Geosci. Electron*., GE-17, IEEE, October 1979.
28. R. M. Wakimoto and R. C. Srivastava (eds.), *Radar and Atmospheric Science: A Collection of Essay in Honor of David Atlas*, Meteorological Monographs, Vol. 30, Boston: AMS, 2003.
29. P. Meischner (ed.), *Weather Radar: Principles and Advanced Applications*, Berlin: Springer-Verlag, 2004.
30. *Preprints and Proceedings of Conferences on Radar Meteorology*, 1947–present, Boston: AMS.
31. *Proceedings of European Conferences on Radar in Meteorology and Hydrology,* 2000–present, Berlin, Germany: Copernicus GmbH.
32. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2001, p. 772.
33. G. Mie, "Beiträge zur Optic trüber Medien, speziell kolloidaler Metallösungen [Contribution to the optics of suspended media, specifically colloidal metal suspensions]," *Ann. Phys.*, vol. 25, pp. 377–445, 1908.
34. J. R. Probert-Jones, "The Radar Equation in Meteorology," *Q. J. R. Meteorol. Soc.*, vol. 88, pp. 485–495, 1962.
35. J. W. Wilson, T. M. Weckwerth, J. Vivekanandan, R. M. Wakimoto, and R. W. Russell, "Boundary layer clear air radar echoes: origin of echoes and accuracy of derived winds," *J. Atmos. Oceanic Technol.*, vol. 11, pp. 1184–1206, 1994.
36. R. W. Russell and J. W. Wilson, "Radar-observed fine lines in the optically clear boundary layer: reflectivity contributions from aerial plankton and its predators," *Boundary Layer Meteorol.,* vol. 82, pp. 235–262, 1997.
37. P. H. Hildebrand, "Iterative correction for attenuation of 5 cm radar in rain," *J. Appl. Meteorol.*, vol. 17, pp. 508–514, 1978.
38. R. H. Allen, D. W. Burgess, and R. J. Donaldson, Jr., "Severe 5-cm radar attenuation of the Wichita Falls storm by intervening precipitation," in *19th Conf. Radar Meteorol.*, AMS, Boston, 1980, pp. 87–89.
39. P. J. Eccles and D. Atlas, "A dual-wavelength radar hail detector," *J. Appl. Meteorol.*, vol. 12, pp. 847–854, 1973.
40. R. E. Carbone, D. Atlas, P. Eccles, R. Fetter, and E. Mueller, "Dual wavelength radar hail detection," *Bull. Amer. Meteor. Soc.*, vol. 54, pp. 921–924, 1973.
41. T. Oguchi, "Electromagnetic wave propagation and scattering in rain and other hydrometeors," *Proc. IEEE*, vol. 71, pp. 1029–1078, 1983.
42. R. J. Donaldson, Jr., "The measurement of cloud liquid-water content by radar," *J. Meteorol*., vol. 12, pp. 238–244, 1955.
43. H. K. Weickmann and H. J. aufm Kampe, "Physical properties of cumulus clouds," *J. Meteorol.*, vol. 10, pp. 204–221, 1953.
44. K. L. S. Gunn and T. W. R. East, "The microwave properties of precipitation particles," *Q. J. R. Meteorol. Soc.*, vol. 80, pp. 522–545, 1954.
45. J. W. Ryde and D. Ryde, *Attenuation of Centimeter Waves by Rain, Hail, Fog, and Clouds*, Wembley, England: General Electric Company, 1945.
46. B. R. Bean and R. Abbott, "Oxygen and water vapor absorption of radio waves in the atmosphere," *Geofts, Pura Appl.*, vol. 37, pp. 127–144, 1957.
47. J. W. Ryde, "The attenuation and radar echoes produced at centimetre wavelengths by various meteorological phenomena," in *Meteorological Factors in Radio Wave Propagation*, London: Physical Society, 1946, pp. 169–188.

48. J. O. Laws and D. A. Parsons, "The relationship of raindrop size to intensity," in *24th Ann. Meet. Trans. Am. Geophys. Union*, 1943, pp. 452–460.
49. R. G. Medhurst, "Rainfall attenuation of centimeter waves: comparison of theory and measurement," *IEEE Trans. Ant. Prop.,* vol. AP-13, pp. 550–564, 1965.
50. R. Uijlenhoet, M. Steiner, and J.A. Smith, "Variability of raindrop size distributions in a squall line and implications for radar rainfall estimation," *J. Hydrometeorol.*, vol. 4, pp. 43–61, 2003.
51. C. R. Burrows and S. S. Attwood, *Radio Wave Propagation, Consolidated Summary Technical Report of the Committee on Propagation, NDRC*, New York: Academic Press, 1949, p. 219.
52. W. J. Humphreys, *Physics of the Air*, New York: McGraw-Hill Book Company, 1940, p. 82.
53. *Glossary of Meteorology*, 2nd Ed., Boston: AMS, 2000, p. 885.
54. D. Atlas and E. Kessler III, "A model atmosphere for widespread precipitation," *Aeronaut. Eng. Rev.*, vol. 16, pp. 69–75, 1957.
55. M. Kerker, M. P. Langleben, and K. L. S. Gunn, "Scattering of microwaves by a melting spherical ice particle," *J. Meteorol*., vol. 8, p. 424, 1951.
56. A. C. Best, *Physics in Meteorology*, London: Sir Isaac Pitman & Sons, Ltd., 1957.
57. J. A. Saxton and H. G. Hopkins, "Some adverse influences of meteorological factors on marine navigational radar," *Proc. IEE (London)*, vol. 98, pt. III, p. 26, 1951.
58. J. N. Chrisman and C. A. Ray, "A first look at the operational (data quality) improvements provided by the Open Radar Data Acquisition (ORDA) system," in *21st Int. Conf. on Infor. Processing Sys. (IIPS) for Meteorol., Oceanog., and Hydrol.*, San Diego, CA, P4R.10, 2005.
59. M. Sachidananda and D. S. Zrnic, "Clutter filtering and spectral moment estimation for doppler weather radars using staggered pulse repetition time (PRT)," *J. Atmos. Ocean. Tech.*, 17, pp. 323–331, 2000.
60. L. B. Jackson, *Digital Filters and Signal Processing*, 2nd Ed., Norwell, MA: Kluwer, 1989.
61. A. D. Siggia and R. E. Passarelli, Jr., "Gaussian model adaptive processing (GMAP) for improved ground clutter cancellation and moment calculation," in *3rd European Conf. on Radar Meteoro.*, Visby, Island of Gotland, Sweden, 2004, pp. 67–73.
62. F. Pasqualucci, B. W. Bartram, R. A. Kropfli, and W. R. Moninger, "A millimeter-wavelength dual-polarization doppler radar for cloud and precipitation studies," *J. Clim. Appl. Meteorol.*, vol. 22, pp. 758–765, 1983.
63. R. Lhermitte, "A 94-GHz doppler radar for cloud observations," *J. Atmos. Ocean. Technol.*, vol. 4, pp. 36–48, 1987.
64. J. H. Richter, "High-resolution tropospheric radar sounding," *Proc. Colloq. Spectra Meteorol. Variables, Radio Sci.*, vol. 4, pp. 1261–1268, 1969.
65. R. J. Keeler, D. S. Zrnic, and C. L. Frush, "Review of range velocity ambiguity mitigation techniques," in *29th Conf. on Radar Meteorol.*, AMS, Montreal, 1999, pp. 158–163.
66. B. G. Laird, "On ambiguity resolution by random phase processing," in *20th Conf. Radar Meteorol.*, Boston, AMS, 1981, p. 327.
67. M. Sachidananda and D. S. Zrnic, "Systematic phase codes for resolving range overlaid signals in a doppler weather radar," *J. Atmos. Oceanic Technol.,* vol. 16, pp. 1351–1363, 1999.
68. J. Pirttilä and M. Lehtinen, "Solving the range-doppler dilemma with the SMPRF pulse code," in *30th Conf. Radar Meteorol.*, Munich, AMS, 2001, pp. 322–324.
69. National Research Council, *Weather Radar Technology beyond Nexrad*, Washington, DC: National Academy Press, 2002.
70. R. J. Keeler, J. Lutz, and J. Vivekanandan, "S-Pol—NCAR's polarimetric doppler research radar," in *Proc. Int. Geosci. Remote Sens. Symp. [IGARSS 2000]*, IEEE, Honolulu, 2000, pp. 1570–1573.
71. H. Liu and V. Chandrasekar, "Classification of hydrometeors based on polarimetric radar measurements: development of fuzzy logic and neuro-fuzzy systems and in-situ verification," *J. Atmos. Ocean. Technol.*, vol. 17, pp. 140–164, 2000.

72. C. J. Kessinger, S. M. Ellis, and J. VanAndel, "The radar echo classifier: a fuzzy logic algorithm for the WSR-88D," presented at *83rd AMS Annual Meeting (3rd AI Conf.)*, P1.6, Long Beach, 2003.
73. D. Atlas, "Radar calibration: some simple approaches," *Bull. Am. Meteorol. Soc.*, vol. 83, pp. 1313–1316, 2002.
74. J. F. Pratt and D. G. Ferraro, "Automated solar gain calibration, preprints," in *24th Conf. Radar Meteorol.*, AMS, Tallahassee, 1989, pp. 619–622.
75. D. Sirmans and B. Urell, "On measuring WSR-88D antenna gain using solar flux," NWS ROC Engineering Branch Report, 2001.
76. E. Gorcucci, J. Scarchilli, and V. Chandrasekar, "Calibration of radars using polarimetric techniques," *IEEE Trans. Geosci. Rem. Sens.*, vol. 30, pp. 853–858, 1992.
77. J. C. Hubbert, V. N. Bringi, and D. Brunkow, "Studies of the polarimetric covariance matrix. Part I: Calibration methodology," *J. Atmos. Ocean. Technol.*, vol. 20, pp. 696–706, 2003.
78. Radar calibration workshop presented at *81st Annual Meeting of the Am. Meteorol. Soc.*, Albuquerque, 2001.
79. R. J. Keeler and R. E. Passarelli, "Signal processing for atmospheric radars," Chapter 20 in *Radar in Meteorology*, Atlas (ed.), Boston: AMS, 1990, pp. 199–229.
80. J. S. Marshall and W. Hitschfeld, "The interpretation of the fluctuating echo for randomly distributed scatterers," Pt. I, *Can. J. Phys.*, vol. 31, pp. 962–994, 1953.
81. P. R. Wallace, "The interpretation of the fluctuating echo for randomly distributed scatterers," Pt. II, *Can. J. Phys.*, vol. 31, pp. 995–1009, 1953.
82. D. S. Zrnic, "Simulation of weather-like doppler spectra and signals," *J. Appl. Meteorol.*, vol. 14, pp. 619–620, 1975.
83. D. S. Zrnic and R. J. Doviak, "Velocity spectra of vortices scanned with a pulse-doppler," *J. Appl. Meteorolo.*, vol. 14, pp. 1531–1539, 1975.
84. W. D. Rummler, "Introduction of a new estimator for velocity spectral parameters," *Tech. Memo MM-68-4121-5*, Bell Telephone Laboratories, Whippany, NJ, 1968.
85. D. S. Zrnić, "Estimating of spectral moments for weather echoes," *IEEE Trans. Geosc. Electron.*, vol. GE-17, pp. 113–128, 1979.
86. A. V. Oppenheim and R. W. Schaefer, *Digital signal processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975.
87. F. Fabry and R. J. Keeler, "Innovative signal utilization and processing," Chapter 8 in *Radar in Atmospheric Science: A Collection of Essays in Honor of David Atlas,* R. Wakimoto and R. Srivastava (eds.), Meteorological Monographs, Vol. 30, Boston: AMS, 2003, pp. 199–214.
88. T. L. Wilfong, D. A. Merritt, R. J. Lataitis, B. L. Weber, D. B. Wuertz, and R. G. Strauch, "Optimal generation of radar wind profiler spectra," *J. Atmos. Ocean. Technol.,* vol. 16, pp. 723–733, 1999.
89. P. H. Hildebrand and R. H. Sekhon, "Objective determination of the noise level in doppler spectra," *J. Appl. Meteorol.*, vol. 13, pp. 808–811, 1974.
90. H. Urkowitz and J. P. Nespor, "Obtaining spectral moments by discrete Fourier transform with noise removal in radar meteorology," *Proc. Int. Geosci. Remote Sens. Symp. [IGARSS-92]*, IEEE, Houston, 1992, pp. 125–127.
91. J. N. Denenberg, R. J. Serafin, and L. C. Peach, "Uncertainties in coherent measurement of the mean frequency and variance of the doppler spectrum from meteorological echoes," in *15th Conf. Radar Meteorol.*, AMS, Boston, 1972, pp. 216–221.
92. R. J. Keeler and C. L. Frush, "Coherent wideband processing of distributed targets," in *Proc. Int. Geosci. and Remote Sensing Symp. [IGARSS-83]*, San Francisco, IEEE/URSI, 1983, pp. 3.1–3.5.
93. R. G. Strauch, "A modulation waveform for short-dwell-time meteorological doppler radars," *J. Atmos. Oceanic Technol.*, vol. 5, pp. 512–520, 1988.
94. R. J. Keeler and C. A. Hwang, "Pulse compression for weather radar," in *IEEE Int. Radar Conf.*, Washington, DC, 1995, pp. 1–7.

95. A. Mudukutore, V. Chandrasekar, and R.J. Keeler, "Pulse compression for weather radars," *IEEE Trans. on Geosci. Rem. Sens.*, vol. 36, pp. 125–142, 1998.
96. F. O'Hora and J. Keeler, "Comparison of pulse compression & whitening transform signal processing," in *4th European Radar Conf.*, Barcelona, 2006, pp. 109–112.
97. E. A. Robinson, "Predictive decomposition of time series with application to seismic exploration," *Geophysics*, vol. 32, pp. 418–484, 1967.
98. R. J. Keeler and L. J. Griffiths, "Acoustic doppler extraction by adaptive linear prediction filtering," *J. Acoust. Soc. Amer.*, vol. 61, pp. 1218–1227, 1977.
99. A. C. Koivunen and A. B. Kostinski, "Feasibility of data whitening to improve performance of weather radar," *J. Appl. Meteorol.,* vol. 38, pp. 741–749, 1999.
100. S. M. Torres and D. S. Zrnic, "Whitening in range to improve weather radar spectral moment estimates. Part 1: formulation and simulation," *J. Atmos. Oceanic Technol.*, vol. 20, pp. 1433–1448, 2003.
101. T. Y. Yu, G. Zhang, A. B. Chalamalasetti, R. J. Doviak, and D. S. Zrnic, "Resolution enhancement technique using range oversampling," *J. Atmos. Ocean. Technol*., vol. 23, pp. 228–240, 2006.
102. A. V. Oppenheim and R. W. Schaefer, *Discrete Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.
103. J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," Geophysics, vol. 37, pp. 375–376, 1972.
104. J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1419, 1969.
105. S. M. Kay, *Modern Spectral Estimation: Theory and Application,* New York: Prentice-Hall, 1988.
106. S. D. Campbell and S. H. Olson, "Recognizing low-altitude wind shear hazards from doppler weather radar: an artificial intelligence approach," *J. Atmos. Ocean. Technol.*, vol. 4, p. 5–18, 1987.
107. A. L. Pazmany, J. B. Mead, S. M. Sekelsky, and D. J. McLaughlin, "Multi-frequency radar estimation of cloud and precipitation properties using an artificial neural network," in *30th Int. Conf. on Radar Meteorol.*, Munich, AMS, pp. 154–156, 2001.
108. D. Atlas, "Advances in radar meteorology," in *Advances in Geophysics*, Vol. 10, New York: Academic Press, 1964.
109. R. Gunn and G. D. Kinzer, "The terminal velocity of fall for water droplets in stagnant Air," *J. Meteorol.*, vol. 6, pp. 243–248, 1949.
110. J. S. Marshall and W. M. K. Palmer, "The distribution of raindrops with size," *J. Meteorol.*, vol. 4, pp. 186–192, 1948.
111. D. C. Blanchard, "Raindrop size distribution in Hawaiian rains," *J. Meteorol.*, vol. 10, pp. 457–473, 1953.
112. D. M. A. Jones, "3 cm and 10 cm wavelength radiation backscatter from rain," in *5th Weather Radar Conf.*, AMS, Boston, 1955, pp. 281–285.
113. K. L. S. Gunn and J. S. Marshall, "The distribution with size of aggregate snowflakes," *J. Meteorol.*, vol. 15, pp. 452–466, 1958.
114. J. W. Wilson and E. A. Brandes, "Radar measurement of rainfall—a summary," *Bull. Am. Meteorol. Soc.*, vol. 60, pp. 1048–1058, 1979.
115. I. Zawadzki, "On radar-raingage comparison," *J. Appl. Meteorol.*, vol. 14, pp. 1430–1436, 1975.
116. I. Zawadzki, "Factors affecting the precision of radar measurements of rain," in *22cd Conf. Radar Meteorol.*, AMS, Boston, 1984, pp. 251–256.
117. J. Joss. and R. Lee, "Application of radar-gauge comparison to operation precipitation profile corrections," *J. Appl. Meteorol.*, vol. 34, pp. 2612–2630, 1995.
118. U. Germann and J. Joss, "Mesobeta profiles to extrapolate radar precipitation measurements above the Alps to ground level," *J. Appl. Meteorol.*, vol. 41, pp. 542–547, 2002.
119. B. Vignal, G. Galli, J. Joss, and U. Germann, "Three methods to determine profiles of reflectivity from volumetric radar data to correct precipitation estimates," *J. Appl. Meteorol.*, vol. 39, pp. 1715–1726, 2000.

120. F. F. Marzano, E. Picciotti, and G. Vulpiani, "Rain field and reflectivity vertical profile reconstruction from C-band radar volumetric data," *IEEE Trans Geosci Rem. Sens.*, vol. 42, pp. 1033–1046, 2004.

121. J. Bridges and J. Feldman, "An attenuation reflectivity technique to determine the drop size distribution of water clouds and rain," *J. Appl. Meteorol.*, vol. 5, pp. 349–357, 1966.

122. D. S. Zrnic and A. Ryzhkov, "Polarimetry for weather surveillance radars," *Bull. Amer. Meteoro. Soc.*, vol. 80, pp. 389–406, 1999.

123. T. A. Seliga and V. N. Bringi, "Potential use of radar differential reflectivity measurements at orthogonal polarizations for measuring precipitation," *J. Appl. Meteorol.*, vol. 15, pp. 69–76, 1976.

124. M. Sachidananda. and D. S. Zrnic, "Rain Rate estimation from differential polarization measurements," *J. Atmos. Ocean. Tech.*, vol. 4, pp. 588–598, 1987.

125. D. N. Moisseev, C. M. H. Unal, H.W. J. Russchenberg, and L.P. Ligthart, "Improved polarimetric calibration of atmospheric radars," *J. Atmos. Ocean. Tech.*, vol. 19, pp. 1968–1977, 2002.

126. A. Ryzhkov and D. Zrnic, "Assessment of rainfall measurement that uses specific differential phase," *J. Appl. Meteorol.* 35, pp. 2080–2090, 1996.

127. G. Zhang, J. Vivekanandan, and E. Brandes, "A method for estimating rain rate and drop size distribution from polarimetric radar measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 39**,** pp. 830–841, 2001.

128. J. Vivekanandan, G. Zhang, S.M. Ellis, D. Rajopadhyaya, and S.K. Avery, "Radar reflectivity calibration using differential propagation phase measurement," *Radio Sci.*, vol. 38, pp. 14–1 to 14–14, 2003.

129. P. Kollias, B. A. Albrecht, and F. Marks, Jr., "Why Mie?," *Bull. Amer. Meteor. Soc.*, vol. 83, pp. 1471–1483, 2002.

130. R. J. Doviak, V. Bringi, A. Ryzhkov, A. Zahrai, and D. Zrnic, "Considerations for polarimetric upgrades to the operational WSR-88D radars," *J. Atmos. Ocean. Technol.*, vol. 17, pp. 257–277, 2000.

131. C. Frush, R. J. Doviak, M. Sachidananda, and D. S. Zrnic, "Application of the SZ phase code to mitigate range-velocity ambiguities in weather radars," *J. Atmos. Ocean. Technol.*, vol. 19, pp. 413–430, 2002.

132. J. C. Hubbert, G. Meymaris, and R. J. Keeler, "Range-velocity mitigation via SZ phase coding with experimental S-band radar data," in *31st Conf. on Radar Meteorol.*, AMS, Seattle, 2003 pp. 727–729.

133. M. Sachidananda and D. S. Zrnic, "Clutter filtering and spectral moment estimation for doppler weather radars using staggered pulse repetition time (PRT)," *J. Atmos. Ocean. Technol*., vol. 17, pp. 323–331, 2000.

134. D. Burgess et al., "Final report on the Joint Doppler Operational Project (JDOP)," 1976–1978, *NOAA Tech. Memo.* ERL NSSL-86, 1979.

135. W. C. Lee and M. M. Bell, "Rapid intensification, eyewall contraction and breakdown of Hurricane Charley (2004) near landfall," *Geophys. Res. Lett.*, vol. 34, L02802, doi:10.1029/2006GL027889, 2007.

136. R. A. Houze Jr., S. S. Chen, W. C. Lee, R. F. Rogers, J. A. Moore, G. J. Stossmeister, J. L. Cetrone, W. Zhao, and M. M. Bell, "The Hurricane Rainband and Intensity Change Experiment (RAINEX): Observations and modeling of Hurricanes Katrina, Ophelia, and Rita (2005)," *Bull. Amer. Meteoro. Soc.*, vol. 87, pp. 1503–1521, 2006.

137. R. J. Donaldson, Jr., "Vortex signature recognition by a doppler radar," *J. Appl. Meteorol.*, vol. 9, pp. 661–670, 1970.

138. J. Wilson and H. P. Roesli, "Use of doppler radar and radar networks in mesoscale analysis and forecasting," *ESA J.*, vol. 9, pp. 125–146, 1985.

139. T. Fujita and F. Caracena, "An analysis of three weather-related aircraft accidents," *Bull. Am. Meteorol. Soc.*, vol. 58, pp. 1164–1181, 1977.

140. T. Fujita, "The downburst," Satellite and Mesometeorology Research Project, Department of the Geophysical Sciences, University of Chicago, 1985.

141. T. Fujita, "The DFW microburst," Satellite and Meteorology Research Project, Department of the Geophysical Sciences, University of Chicago, 1986.

142. J. McCarthy and R. Serafin, "The microburst: hazard to aviation," *Weatherwise*, vol. 37, pp. 120–127, 1984.
143. R. D. Roberts and J. W. Wilson, "A proposed microburst nowcasting procedure using single doppler radar," *J. Appl. Meteorol.,* vol. 28, pp. 285–303, 1989.
144. K. Aydin, T. A. Seliga, and V. Balaji, "Remote sensing of hail with dual linear polarization radar," *J. Clim. Appl. Meteorol.*, vol. 25, pp. 1475–1484, 1986.
145. A. J. Illingworth, J. W. F. Goddard, and S. M. Cherry, "Detection of hail by dual polarization radar," *Nature*, vol. 320, pp. 431–433, 1986.
146. R. M. Lhermitte and D. Atlas, "Precipitation motion by pulse doppler radar," in *9th Weather Radar Conf.*, AMS, Boston, 1961, pp. 218–223.
147. K. A. Browning and R. Wexler, "A determination of kinematic properties of a wind field using doppler radar," *J. Appl. Meteorol*., vol. 7, pp. 105–113, 1968.
148. J. D. Tuttle and G. B. Foote, "Determination of the boundary layer airflow from a single doppler radar," *J. Atmos. Ocean. Technol.*, vol. 7, pp. 218–232, 1990.
149. J. W. Wilson and W. E. Schreiber, "Initiation of convective storms at radar-observed boundary layer convergence lines," *Mon. Weather Rev.*, vol. 114, pp. 2516–2536, 1986.
150. T. W. Weckwerth, C. R. Pettet, F. Fabry, S. J. Park, M. A. LeMone, and J. W. Wilson, "Radar refractivity retrieval: Validation and application to short-term forecasting," *J. Appl. Meteoro.*, vol. 44, pp. 285–300, 2005.
151. M. Hall (ed.), "Special papers: Multiple parameter radar measurements of precipitation," *Radio Sci.*, vol. 19, 1984.
152. R. M. Lhermitte, "Dual-doppler radar observations of convective storm circulations," in *14th Conf. Radar Meteorol.*, AMS, Boston, 1970, pp. 139–144.
153. D. J. McLaughlin, V. Chandrasekar, K. Droegemeier, S. Frasier, J. Kurose, F. Junyent, B. Philips, S. Cruz-Pol, and J. Colom, "Distributed Collaborative Adaptive Sensing (DCAS) for improved detection, understanding and prediction of atmospheric hazards," presented at *85th AMS Annual Meeting*, San Diego, AMS, 2005.
154. F. Junyent, V. Chandrasekar, D. Brunkow, P. C. Kennedy, and D. J. McLaughlin, "Validation of first generation CASA radars with CSU-CHILL," presented at *32cd Conf. Radar Meteorol.*, P10R4, AMS, Albuquerque, 2006.
155. R. E. Carbone, M. Carpenter, and C. Burghart, "Doppler radar sampling limitations in convective storms," *J. Atmos. Ocean. Technol.*, vol. 2, pp. 358–361, 1985.
156. M. Brook and P. Krehbiel, "A fast-scanning meteorological radar," in *16th Conf. Radar Meteorol.*, AMS, Boston, 1975, pp. 26–31.
157. R. J. Keeler and C. L. Frush, "Rapid-scan doppler radar development considerations, Part II: technology assessment," in *21st Conf. Radar Meteorol.* AMS, Boston, 1983, pp. 284–290.
158. P. L. Smith, "Applications of radar to meteorological operations and research," *IEEE Proc.*, vol. 62, pp. 724–725, 1974.
159. C. L. Holloway and R. J. Keeler, "Rapid scan doppler radar: the antenna issues," in *26th Conf. on Radar Meteorol.*, AMS, Norman, 1993, pp. 393–395.
160. L. Josefsson, "Phased array antenna technology for weather radar applications," in *25th Conf. on Radar Meteorol.*, AMS, Paris, 1991, pp. 752–755.
161. R. J. Keeler, "Weather radars of the 21st century: a technology perspective," in *28th Conf. on Radar Meteorol.*, Austin, AMS, 1997, pp. 309–310.
162. P. Meischner, C. Collier, A. Illingworth, J. Joss, and W. Randeu, "Advanced weather radar systems in Europe: The COST-75 action," *Bull. Amer. Meteoro. Soc.*, vol. 78, pp. 1411–1430, 1997.
163. E. Brookner (ed.), *Practical Phased Array Antenna Systems,* Norwich, MA: Artech House, 1991.
164. J. Wurman and M. Randall, "An inexpensive, mobile, rapid scan radar," in *30th Int. Conf. on Radar Meteorol.*, Munich, AMS, 2001, pp. 98–100.
165. J. W. Rogers, L. Buckler, A. C. Harris, M. Keehan, and C. J. Tidwell, "History of the Terminal Area Surveillance System (TASS)," in *28th Conf. Radar Meterol.,* AMS, Austin, 1997, pp. 157–158.

166. W. Benner, W. G. Torok, N. Gordner-Kalani, M. Batista-Carver, and T. Lee, "MPAR program overview and status," presented at *23rd Int. Conf. on Interact. Info. Proc. Sys. (IIPS)*, AMS, San Antonio, 2007.

167. T. Maese, J. Melody, S. Katz, M. Olster, W. Sabin, A. Freedman, and H. Owen, "Dual-use shipborne phased array radar technology and tactical environmental sensing," in *Proc. IEEE National Radar Conf.*, Atlanta, 2001, pp. 7–12.

168. D. E. Forsyth, K. J. Kimpel, D. S. Zrnic, S. Sandgathe, R. Ferek, J. F. Heimmer, T. McNellis, J. E. Crain, A. M. Shapiro, J. D. Belville, and W. Benner, "The national weather radar testbed (phased array)," presented at *18th Int. Conf. on Interact. Info. Proc. Sys. (IIPS)*, AMS, Orlando, 2002.

169. R. L. Trotter, "Design considerations for the NOAA airborne meteorological radar and data system," in *18th Conf. on Radar Meteorol.*, AMS, Atlanta, 1978, pp. 405–408.

170. H. B. Bluestein and R. M. Wakimoto, "Mobile radar observation of severe convective storms," Chapter 5 in *Radar in Atmospheric Science: A collection of essays in honor of David Atlas*, R. Wakimoto and R. Srivastava (eds.), Meteorological Monograph, Vol. 30, Boston: AMS, 2003, pp. 105–136.

171. P. H. Hildebrand, C. A. Walther, C. L. Frush, J. Testud, and F. Baudin, "The ELDORA/ASTRAIA airborne doppler weather radar: goals, design and first field test," *Proc. IEEE*, vol. 12, pp. 1873–1890, 1994.

172. G. M. Heymsfield et al., "The EDOP radar system on the high altitude NASA ER-2 aircraft," *J. Atmos, Oceanic Technol*., vol. 13, pp. 795–809, 1996.

173. L. Li, G. M. Heymsfield, P. E. Racette, L. Tian, and E. Zenker, "A 94-GHz cloud radar system on a NASA high-altitude ER-2 aircraft," *J. Atmos. Oceanic Technol.*, vol. 21, pp. 1378–1388, 2004.

174. D. P. Jorgensen, T. R. Shepherd, and A. S. Goldstein, "A dual-pulse repetition frequency scheme for mitigating velocity ambiguities of the NOAA P-3 airborne doppler radar," *J. Atmos. Oceanic Technol.*, vol. 17, pp. 585–594, 2000.

175. S. L. Durden, E. Im, F. K. Li, W. Ricketts, A. Tanner, and W. Wilson, "ARMAR: An airborne rain mapping radar," *J. Atmos. Oceanic Technol.*, vol. 11, pp. 727–737, 1994.

176. A. Pazmany, R. McIntosh, R. Kelly, and G. Vali, "An airborne 95 GHz dual-polarized radar for cloud studies," *IEEE Trans. Geosci. Remote Sens*, vol. 32, pp. 731–739, 1994.

177. G. Farquharson, E. Loew, W. C. Lee, and J. Vivekanandan, "A new high-altitude airborne millimeter-wave radar for atmospheric research," presented at *Proc. Int. Geosci. Remote Sens. Symp. [IGARSS 2007]*, IEEE, Barcelona, 2007.

178. M. Wolde and A. Pazmany, "NRC Dual-frequency airborne radar for atmospheric research," presented at *32cd Conf. Radar Meteorol.*, P1R.9, Albuquerque, 2005.

179. T. Kozu et al., "Development of precipitation radar on-board the Tropical Rainfall Measuring Mission (TRMM) satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 102–116, 2001.

180. E. Im et al., "Second-generation precipitation radar (PR-2)," Final Rep. JPL D-22997, NASA Earth Science Instrument Incubator Program, JPL, Calif. Inst. Tech., Pasadena, CA, 2002.

181. R. Meneghini and D. Atlas, "Simultaneous ocean cross-section and rainfall measurements from space with a nadir-looking radar," *J. Atmos. Ocean. Technol.*, vol. 3, pp. 400–413, 1986.

182. L. Liang and R. Meneghini, "A study of air/space-borne dual-wavelength radar for estimation of rain profiles," *Advances in Atmos. Sci,* vol. 22, pp. 841–851, 2005.

183. G. L. Stephens, D. G. Vane, R. J. Boain, G. G. Mace, K. Sassen, Z. Wang, A. J. Illingworth, E. J. O'Connor, W. B. Rossow, S. L. Durden, S. D. Miller, R. T. Austin, A. Benedetti, C. Mitrescu, and the CloudSat Science Team, "The CloudSat mission and the A-train: a new dimension of space-based observations of clouds and precipitation," *Bull. Am. Meteorol. Soc.*, vol. 83, pp. 1771–1790, 2002.

184. E. E. Gossard and R. G. Strauch, *Radar Observations of Clear Air and Clouds*, Amsterdam: Elsevier, 1983.

185. S. Kato T. Tsuda, M. Yamamato, T. Sato, and S. Fukao, "First results obtained with a middle and upper atmosphere (MU) radar," *J. Atmos. Terr. Phys.*, vol. 48, pp. 1259–1267, 1986.

186. S. G. Benjamin, B. E. Schwartz, E. J. Szoke, and S. E. Koch, "The value of wind profiler data in U.S. weather forecasting," *Bull. Amer. Meteor. Soc.*, vol. 85, pp. 1871–1886, 2004.

187. R. E. Carbone, R. Strauch, and G. M. Heymsfield, "Simulation of wind profilers in distributed conditions," in *23rd Conf. Radar Meteorol.*, vol. I, AMS, Boston, 1986, pp. 44–47.

188. R. G. Strauch, B. L. Weber, A. S. Frisch, C. G. Little, D. A. Merritt, K. P. Moran, and D. C. Welsh, "The precision and relative accuracy of profiler wind measurements," *J. Atmos. Ocean. Technol.*, vol. 4, pp. 563–571, 1987.

189. *Proceedings of International Symposia on Tropospheric Profiling*, 1998–present.

190. R. J. Doviak, R. J. Lataitis, and C. L. Holloway, "Cross correlations and cross spectra for spaced antenna wind profilers," *Radio Sci.*, vol. 31, pp. 157–180, 1996.

191. J. S. Van Baelen and A. D. Richmond, "Radar interferometry technique: Three-dimensional wind measurement theory," *Radio Sci.,* vol. 26, pp. 1209–1218, 1991.

192. G. Zhang and R. J. Doviak, "Spaced-antenna interferometry to measure crossbeam wind, shear and turbulence: Theory and formulation," *J. Atmos. Ocean. Technol.*, vol. 24, pp. 791–805, 2007.

Chapter 20
HF Over-the-Horizon Radar

James M. Headrick

Naval Research Laboratory (retired)

Stuart J. Anderson

Australian Defence Science and Technology Organisation

20.1 INTRODUCTION

Beyond-the-horizon detection of terrestrial targets at ranges of thousands of kilometers can be achieved by radars operating in the high-frequency (HF) band (3 to 30 MHz). This very long range coverage is obtained by using *skywave* propagation, that is, reflecting the radar signals from the ionosphere. HF ground wave (surface wave) propagation over the sea has been used for intermediate but still over-the-horizon distances, up to several hundred kilometers. Occasionally HF radar systems also find applications in line-of-sight applications at short range. This chapter focuses predominantly on skywave radar, though much of the discussion applies equally to surface wave radar. Rather than amalgamate discussion of the two radar configurations throughout the chapter, the distinctive features of HF surface wave radar are treated separately in an appendix at the end of the chapter.

In one sense. the development of HF skywave radar can be traced back to the 1920s, when skywave echoes were identified,[1] but the first HF radar systems were not deployed until the 1950s.[2] Since then, skywave radar has evolved to address applications such as the detection and tracking of aircraft, ballistic and cruise missiles, and ships.[3–15] In addition to detecting "skin" echoes from targets of interest, HF radar is useful for observing various forms of high-altitude atmospheric ionization, both natural, including those associated with aurorae and meteors, and artificial, including the interaction of spacecraft and ballistic missiles with the ionospheric plasma.[16–19] Further, the wavelengths used are of the same order as ocean surface gravity waves, and this correspondence can be exploited to provide information on the wave directional spectrum, ocean currents, and, by inference, surface winds.[5] Indeed, scattering from the sea can often be employed as a radar cross section (RCS) amplitude reference and is a widely used diagnostic tool. The narrow-band waveforms employed, the low frequencies, and the nature of the transmission path make the spatial resolution coarse when compared with higher-frequency radars, but the doppler resolution can be exceedingly fine. The magnitude and doppler distribution of the echoes from the distant Earth's surface (often termed *backscatter*, though that term should be reserved for monostatic radars) are major factors in setting system dynamic range, spectral purity, and signal processing requirements.

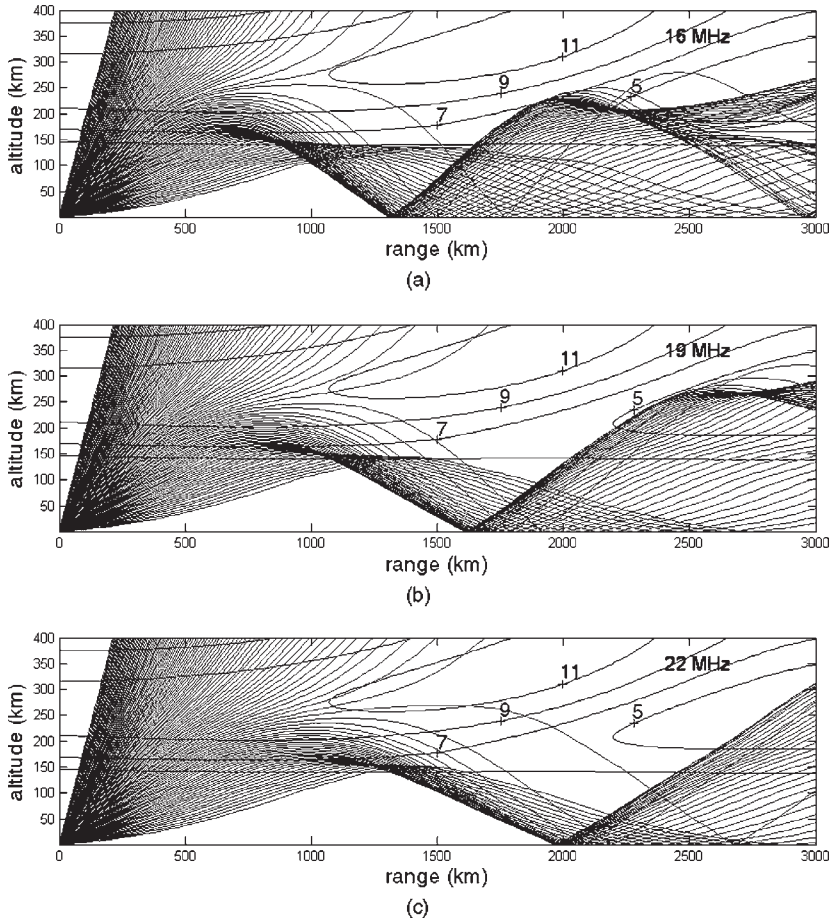


FIGURE 20.1 Ray-tracing through a model ionosphere, showing the variation of the radar footprint with carrier frequency. The contours map the plasma frequency or electron density.

For effective radar operation, environmental parameters that affect radar performance need to be determined in real time, defined as that interval in which there are no large-scale changes in the ionosphere. Typically, this is of the order of 10–30 minutes. Transmission-path information is generally derived from adjunct vertical and oblique sounders as well as by using the radar itself as a sounder. An ionospheric electron density model complex enough to enable adequate sounding interpretation is required. Ionospheric or transmission path statistical climatologies and forecasts are necessary for radar design and for development of site-specific models. In addition, other users in the HF spectrum must be observed continuously and operating frequencies selected to avoid interference.

The essential features of skywave propagation can be seen in Figure 20.1. The ionosphere, being an ionized gas with free electrons, will reflect all radar signals when the radar frequency is less than the *plasma frequency* given by

$$f_P \cong 9 \times 10^{-6} \sqrt{N_e} \tag{20.1}$$

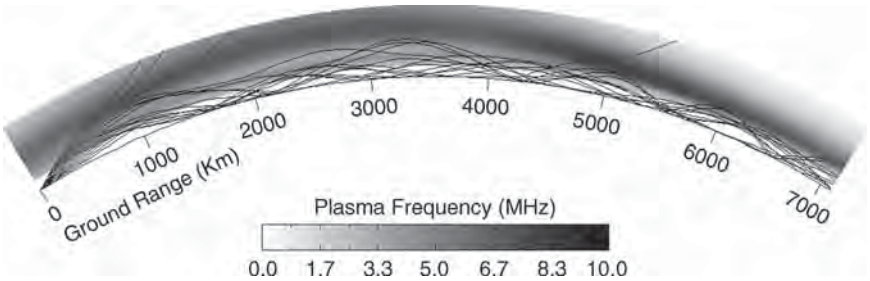


FIGURE 20.2 Numerically computed rays illustrating multiple hop propagation across the equatorial zone. The elevated peak electron density near the equator at ~2500 km is the Appleton anomaly.

where f_P is in MHz and N_e is in electrons per cubic meter. For a given elevation angle α, if the radar frequency exceeds $f_p / \sin\alpha$ at the height of maximum ionization, rays launched at higher elevation angles will escape, resulting in a so-called *skip* or *dead* zone within which the Earth's surface is not illuminated. Beyond this skip zone, energy is returned to the Earth, reaching a maximum range when the rays leaving the antenna are horizontal. Useful range coverage will lie between these limits, which define the *one-hop zone*.

As shown in Figure 20.2, a multiplicity of hops may exist, and energy may even circle the earth; clutter echoes from these long ranges can seriously degrade radar performance. A comparison of Figures 20.1*a,* 20.1*b,* and 20.1*c* reveals that different range extents are illuminated by using different operating frequencies, with longer starting ranges requiring higher frequencies. In the examples shown, 16 MHz illuminates ranges from 1300 to 3000 km, whereas 19 MHz illuminates ranges from 1650 to 2750 km whereas 22 MHz illuminates from 1950 to 2750 km. Hence, the farthest edge of the footprint need not increase with frequency, depending on the prevailing ionospheric conditions. In this example, a single ionospheric layer is considered. Normally, there are two or three distinct layers such that signals may partially penetrate the lower layers to be reflected by a higher layer. As a consequence, the relationship between the range to a target and the measured echo time delay becomes multivalued, with unknown parameters such as layer heights that must be estimated by various techniques as described later in this chapter.

To illustrate the operating principles of a skywave radar, Figure 20.3 presents a map showing multiple surveillance tasks assigned to a hypothetical radar with 360° azimuthal coverage. Five sectors of current surveillance activity are shown, each addressing a particular mission as indicated, each mission consisting of a number of *dwell interrogation regions (DIRs)*. The electronically steered radar transmit beam steps through these DIRs, in some assigned sequence, illuminating each DIR with an appropriate waveform for a corresponding interval during which the receiving system acquires a coherent time series of echo samples. The coherent integration time (CIT) depends on the type of observation but is almost always in the range of 1–100 seconds. Each transmitter footprint is analyzed by forming simultaneous contiguous receive beams, on the order of 1/2° wide at 15 MHz in the case of the U.S. Navy ROTHR and the Australian Jindalee and JORN radars, which equates to a 10 km cross-range resolution at a 1200 km range. The task-specific requirements on DIR revisit rates determine the sequencing of the DIR interrogations and, of course, set the limit on how many tasks can be addressed.

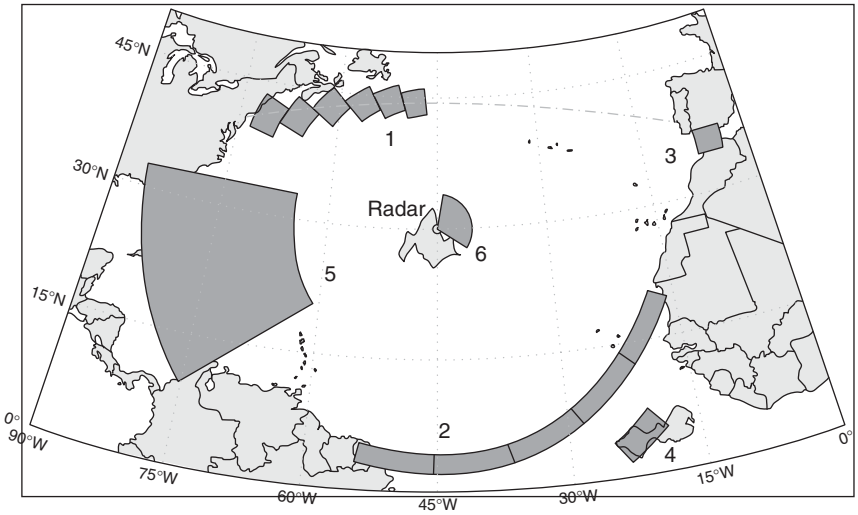


FIGURE 20.3 Some coverage and tasking options for hypothetical HF radars on a (fictitious) mid-Atlantic island. The coverage regions numbered 1–5 correspond to skywave radar missions as follows: 1, air route monitoring; 2, barrier surveillance; 3, strategic waterway monitoring; 4, ballistic missile launch detection; and 5, remote sensing and hurricane tracking. The sector designated 6 is representative of HF surface wave radar coverage.

Consider the set of tasks and DIRs shown in Figure 20.3. Task 1 would require only short CITs, 1–2 s, say, if the aircraft is assumed to be large, and would need revisits perhaps every minute, as the flight is not expected to maneuver so the track will be well-behaved and observed position errors would be due to ionospheric fluctuations. Only the single DIR containing the aircraft need be interrogated. Task 2, barrier surveillance, if concerned with ship traffic, can afford to relax the revisit time to tens of minutes as ships move so slowly, but in order to achieve detection, long CITs of 20–30 s are required to separate the ship echoes from the sea clutter in the doppler spectrum. Even so, the radar can step across the barrier arc, with plenty of time to interleave a Task 1 dwell between addressing successive DIRs in Task 2. If the barrier Task 2 were concerned with aircraft, with a CIT of 1–2 s, the revisits would need to be frequent enough to achieve the desired probability of detection before the aircraft had time to cross the barrier. One approach is to widen the barrier by processing more range cells or by reducing the waveform bandwidth and allowing the range cells to grow wider, but this may fail if the ionosphere does not support propagation over the increased range depth. Another possibility is to increase the revisit interval over a subset of the DIRs. The relative priority of Tasks 1 and 2 would also need to be taken into account.

Task 4 (ballistic missile launch detection) would require much more frequent surveillance if a launch were believed to be imminent, or else the missile might escape from the radar footprint without being detected. Assume a CIT of 5 seconds and a tolerable revisit interval of 10 s. Tasks 1 and 4 could be interleaved, for example, by carrying out five dwells of Task 4, and then one of Task 1, and then repeating this pattern. Task 3 is concerned with slow-moving ships, so the revisit interval could be tens

of minutes, but a long CIT of perhaps 20–30 s is required in order to separate the ship echoes from the sea clutter in the doppler spectrum. During this time, Task 4's requirements would be violated. Tasks 3 and 4, as defined, are incompatible.

Task 5 (remote sensing of ocean conditions) would have a much lower priority as it deals with slowly varying phenomena, so it would need only occasional revisiting. The smaller sector shown as Task 6, extending to a range of 400 km, is representative of the coverage of an HF surface wave radar, as used for protecting the approaches to a port, for example.

This example is typical of the scheduling and resource allocation problem that is central to HF skywave radar operations. Compromises are often unavoidable. The situation is further complicated by the continuous variation of ionospheric propagation conditions, which must be monitored and used to guide tasking, as when bad ionospheric "weather" is likely to preclude certain missions.

Almost invariably, the sectors and, indeed, the individual DIRs within a sector will require different carrier frequencies, even for a given range, as the ionosphere may vary substantially across the full range and azimuth extent of the active coverage. The defining characteristic of HF skywave radar is that the radar operator must select frequencies that are optimum for the various tasks and adapt these frequencies to the ever-changing ionosphere.

This chapter sets out to explain the principal features of HF skywave radar as it is presently implemented, emphasizing the physical considerations that govern system design and performance.

20.2 THE RADAR EQUATION

A form of the radar equation, Eq. 20.2, can be used to point to aspects of HF radar that are significantly different from radars that use higher frequencies. These differences include adaptation to environment, frequency and waveform selection, radar cross section, path losses, multipath effects, noise, interference, antenna gain, spatial resolution, and clutter. For the case of noise-limited detection, the radar equation takes the form

$$\frac{S}{N} = \frac{P_{\text{av}} G_t G_r T \lambda^2 \sigma F_p}{(4\pi)^3 L_p L_s N_0 R^4} \tag{20.2}$$

where

S/N = output signal-to-noise ratio
P_{av} = average transmitted power
G_t = transmitter antenna gain
G_r = receiver antenna gain
T = coherent processing time
λ = wavelength
σ = target radar cross section
F_p = propagation-path factor
N_0 = noise power per unit bandwidth
L_p, L_s = transmission-path and system losses
R = distance between radar and target

These parameters are explained briefly as follows:

Antennas, G_t and G_r: A common convention for HF-band radars is to include the effect of the ground in the antenna performance characterization, and that convention will be used here. For example, a half-wave dipole in free space has a maximum gain over an isotrope of 2.15 dB. If that antenna is oriented vertically, just above but not touching a perfectly conducting earth, its maximum gain will be increased by a factor of 4, or 6 dB, to 8.15 dB at 0° elevation angle. Because the Earth is never perfect, its conductivity and dielectric constant are factors in determining antenna performance. The electrical properties of the Earth are a much stronger factor for vertical polarization than for horizontal; however, terrain features and surface roughness are important for both polarizations.

Coherent processing time, T: For ranges beyond the skip distance, HF radar returns almost invariably contain Earth backscatter at the same ranges as targets. Doppler processing is used to separate targets from Earth backscatter; hence, coherent samples are acquired over an interval *T,* which may exceed 100 s, though it is usually in the range of 1–20 s.

Wavelength (λ): The wavelength or operating frequency must be selected so that the transmitted energy is reflected by the ionosphere to illuminate the desired area of the Earth. The spectrum of the emissions must be constrained not to interfere with other users. Because both the ionosphere and the HF-band occupancy distributions are time-varying parameters, adaptive radar management is required.

Radar cross section (RCS), σ: The radar cross section of conventional targets will generally be a function of frequency, polarization, and aspect angle, but at HF, the target dimensions are typically of the same order as the wavelength, so scattering behavior is different from that observed at higher frequencies. Scattering also occurs from the environment—clutter—so models of the scattering coefficient per unit area of land and ocean surfaces, or per unit volume of the turbulent ionosphere, are used to provide effective RCS values for Eq. 20.2 when those natural scatterers are the "targets" of interest. Thus, for the RCS of Earth clutter, the normalized surface scattering coefficient σ° is multiplied by the resolution cell size A. The important resolution cell size factors, viz. receiver antenna beamwidth and spectral bandwidth, are not explicitly contained in Eq. 20.2 (Clutter frequently sets the limit to target detectability; in which case, it is the signal-to-clutter ratio rather than the signal-to-noise ratio that is of interest. Accordingly, a different form of the radar equation must then be employed.)

Propagation factors (F_p): Several propagation phenomena, including Faraday polarization rotation, ground-reflection multipath, multiple hop propagation, and ionospheric focusing may need to be included in the equation, depending on the scenario of interest. Faraday rotation refers to the variation of the polarization of the signal incident on the target as a function of time and distance, arising from its propagation through the magnetized ionospheric plasma; linearly polarized transmitted signals often arrive in the target zone with a rotated axis of polarization but still essentially linearly polarized. Since many targets have RCS that vary with polarization, an important result is that the most favorable polarization will illuminate the target recurrently. The spatial scale of the polarization "fringes" in the radar footprint is typically in the range 10–100 km, and the change of frequency needed to rotate the plane of polarization by 90° at a given location in the radar footprint (the polarization bandwidth) is of the order of 100 kHz, so differential

effects may be significant. Of course, polarization will also fluctuate at the receiving antenna as a consequence of the time-varying return path.

Noise (N_o): For radars operating in the HF band, the receiver internal noise is almost always less than the external noise.

Losses (L_p, L_s): The loss term L_p contains the two-way losses along the path traversed, including ionospheric absorption and ground-reflection losses; L_s represents any radar system losses. Ionospheric losses, while predicted on a statistical basis, can constitute a major unknown in real-time radar operation.

Range (R): The range in Eq. 20.2 is the "slant range," that is, the length of the skywave path between target and radar, not the distance as measured along the Earth's surface. The ionospheric reflection height needs to be used to convert this slant range to great-circle ground distance. The apparent range to a particular target may take on more than one value since multiple paths may exist.

With these interpretations, the radar equation (Eq. 20.2) can be used to model the performance of skywave radars when noise, not clutter, is setting the limit to target detectability, noting that the complexity and statistical nature of ionospheric propagation and the external noise environment means that it is often necessary to apply the equation to probability distributions, not scalar values of the parameters. When the target's velocity places its doppler shift beyond those of any clutter returns, the noise-limited model is appropriate, but there are two important situations where this is not the case. The first is ship detection, where the intrinsic doppler spread of sea clutter routinely extends beyond the typical doppler shifts of most ship echoes. The second case is the phenomenon of spread doppler clutter, which arises from plasma instabilities and turbulence, especially post-sunset and at high and low latitudes. The equivalent velocity of this type of clutter can extend to hundreds of meters per second, masking even fast aircraft returns. These topics are addressed in Section 20.8. Dealing with clutter-limited situations is a vital part of the radar designer's task, requiring a detailed understanding of the phenomena and their distributions.

20.3 FACTORS INFLUENCING SKYWAVE RADAR DESIGN

Principal Differences Between HF and Microwave Radar. Before analyzing HF skywave radar systems in detail in the sections that follow and describing those properties of the environment that impact their design and performance, it is instructive to summarize the principle differences between skywave radar and conventional microwave radar. This provides a cautionary reminder not to extrapolate too readily from the familiar characteristics of the microwave domain to the HF band.

HF skywave radars operate at ranges about an order of magnitude greater than microwave long-range air surveillance radars. The HF radar wavelength is hundreds of times greater, so the antennas are proportionately larger, as much as two or three kilometers in length if they are to see ships, but considerably less if only aircraft are to be detected. Transmitter average power might be on the order of several hundred kilowatts for a skywave radar, but on the order of a few kilowatts for a microwave ATC radar. The observation time (CIT) for skywave systems can range from one to many tens of seconds, but is on the order of tens of milliseconds for microwave radar.

The long observation time for a skywave radar is needed to obtain the necessary echo signal energy to ensure reliable target detection as well as to obtain the long integration times needed for effective doppler processing. The ionosphere has a dominating influence on skywave radar, whereas the normal atmosphere has very little effect on microwave radar. The skywave radar's frequency and other parameters are driven primarily by the need to propagate via the ionosphere. Constraints imposed by propagation and the availability of unoccupied frequency channels dictate that the range resolution of HF radars is not nearly as good as that of microwave systems. The HF transmitter has to maintain stringent control of its radiated signal spectrum so as to avoid interference to other users of the HF spectrum. (A similar consideration applies to microwave radars, but not to the extent that it does at HF. Military microwave radars like to have available a wide spectral width for purposes of electronic protection and to extract more detailed target information, but the increasing demands of civilian wireless services have reduced the available spectrum available to microwave radars to the point where it is limiting performance.)

The sensitivity of microwave radars is limited by receiver noise, but the sensitivity of HF radars is limited by external noise that enters the receiver through the antenna. This external noise is due not only to natural mechanisms such as thunderstorms but to the signals from the many HF transmitters throughout the world. Both microwave and HF radars can be limited by the large echo signals from land or sea, though in HF skywave radars, the problem is particularly severe. Doppler processing is essential under such conditions. For some aircraft, the radar cross section at HF is significantly larger than that at microwaves. Many HF OTH radars utilize an FM-CW waveform, so widely separated sites are needed to minimize the leakage of the transmitter into the receiver. Microwave radars have used FM-CW waveforms in the past, but in most cases, these have been replaced by waveforms that don't require separate transmitter and receiver sites.

Table 20.1 presents a comparison of some key radar parameters for representative radar systems of each type, and contrasts the ways in which the respective modes of propagation, scattering, noise, and deployment constrain the form and function of radar systems.

Implications for Skywave Radar Design. From an examination of Table 20.1, it is clear that a skywave radar is not simply a microwave radar scaled up in size by a factor of ~ 1000, that is, in proportion to the wavelength. Referring to the radar equation, the R^{-4} loss term means that, for a representative tenfold increase in detection range of HF over microwave, 40 dB extra range loss accumulates. This cannot all be recovered by radiating more power and increasing antenna gain, for practical engineering reasons and cost, apart from constraints imposed by ionospheric propagation. Coherent (doppler) processing provides the necessary processing gain, but the processing times required—from 1 to 100 s—mean that revisit rates when conducting surveillance over an extended region will fall below acceptable levels unless multiple receive beams are processed in parallel. Designers have explored different trade-offs, in most cases, converging on schemes in which 15–30 receive beams are formed within a broader transmitted beam footprint. The reduced transmitting antenna gain may be compensated by increasing power or coherent processing time.

Even under benign conditions, the ionosphere seldom supports highly coherent propagation over bandwidths greater than ~200 kHz, even when clear channels wide enough to accommodate such waveforms are available, which is infrequently. More typically, clear channels range from 10–50 kHz, so the waveform bandwidth is normally chosen to lie in this range. The corresponding range resolutions extend

from 3–15 km. The cross-range resolution cell dimension $L(R)$ at range R is given by $L \approx \frac{R\lambda}{D}$ where D is the receiving array aperture and λ is the radar wavelength.

TABLE 20.1 Key Differences Between Microwave Radar and HF Skywave Radar (The parameter values quoted here are intended to be broadly representative, rather than an attempt to span all known systems.)

	I. Microwave radar	II. HF skywave radar
Type chosen for comparison	Long-range air traffic control such as ARSR-3	Joint aircraft and ship detection, two-site radar such as Jindalee, JORN, and ROTHR
Major antenna dimension (m)	10–15	1000–3000
Average transmit power (kW)	4	400
Antenna radiation pattern	Fully defined by antenna construction	Strongly influenced by ground properties around antenna
Typical operating range (km)	280–450	1000–4000
Minimum range (km)		1000
Range resolution (m)	300	1500–15000
Propagation medium	• Homogeneous or stratified • Nondispersive • Isotropic • Stable • Linear • Nearly constant	• Structured horizontally and vertically, both deterministically and randomly, on many scales • Frequency dispersive • Anisotropic (magnetoionic) • Highly dynamic • Weakly nonlinear • Varies dramatically with time of day, season, etc.
Radar signal propagation path	• Line-of-sight • Usually unique; may have simple ground-reflection multipath • Relatively stable	• Reflected from ionosphere • Multiple paths, resulting in multiple echoes from a single target, at different apparent ranges, bearings, elevation angles, and doppler shifts • Unstable • Poorly known, must be inferred from ancillary sounding systems
Dominant propagation effects on the radar signal	• Multipath interference between direct and surface-reflected signals	• Attenuation • Focusing and defocusing • Polarization transformation • Phase modulation • Wavefront distortion
Target scattering regime	Optical (high frequency), i.e., target size $\gg$ radar wavelength	Rayleigh – resonance, i.e., target size $<$ or $\sim$ radar wavelength
Clutter	• Can be serious, especially at short ranges • Minimized by narrow beam, short pulse, and doppler processing	Look-down viewing geometry inevitably results in strong ground echoes at the same range as the target, typically 20–80 dB stronger than the target echo
Doppler processing	Widely used for detection of moving targets in clutter	Essential to separate moving targets from strong clutter returns

TABLE 20.1 Key Differences Between Microwave Radar and HF Skywave Radar (The parameter values quoted here are intended to be broadly representative, rather than an attempt to span all known systems.) (*Continued*)

	I. Microwave radar	II. HF skywave radar
Frequency constraints	Can be serious because of the need for wideband radar systems and by competition for the microwave frequency spectrum by communications and other electromagnetic services	• Bounded above by the statistical availability of skywave propagation to ranges of interest • Bounded below by spectrum availability, antenna size, and the rapid fall-off in target RCS • Must not interfere with other users in the crowded HF spectrum, thus limiting choice of frequency and bandwidth • Must adapt continually to the changing ionosphere so as to maintain illumination of current target region
Noise floor dominated by	Internal receiver noise (thermal, etc.)	Sources (atmospheric, galactic, anthropogenic, etc.)
Siting constraints	• Unobstructed, elevated sites preferred	• Receive array site must be EM quiet, generally rural, to avoid city and industrial noise at HF frequencies • Huge arrays require flat, open spaces to minimize topographic effects on beam patterns • If a bistatic or two-site quasi-monostatic design is adopted, it needs two sites with adequate separation (~100 km) and the correct geographical relationship relative to the coverage arc • Location on the Earth must be such that auroral and equatorial spread doppler echoes don't mask targets

Hence, to achieve $L = 10$ km at 1200 km range when the frequency is 15 MHz, and thus a resolution cell not too eccentric in shape, requires an array aperture of about 2400 m. Feeding arrays of this size can require hundreds of kilometers of cables or, in some systems, fiber optics, and raises challenging problems in array calibration. The transmitting array aperture, assuming ≈ 20 receive beams to be fitted into the transmit beam, need be only ~ 120 m at this frequency.

It follows from the radar equation and the considerations just discussed that transmit power in the range of 200–1000 kW is generally appropriate for detecting small aircraft targets over a wide range of conditions. In order to radiate power efficiently, the transmitting antenna elements tend to be large, resonant structures; for example, the vertical log-periodic antennas used in the JORN system are up to 43 m tall. In contrast, the choice of receiving antenna element has traditionally been based on the

precept that efficiency is of low importance for HF reception because external noise is almost always far stronger than internal noise. A more efficient receiving element accepts more signal power but equally more external noise so, *prima facie*, nothing is gained in terms of the signal-to-noise ratio. Cost is thus reduced by employing a small receiving antenna element. Heights of 4–6 m are common.

The look-down geometry of skywave illumination results in strong clutter returns in the same range cell as the target, demanding a high dynamic range able to support clutter-to-target energy ratios in excess of 80 dB. This places severe demands on the spectral purity and dynamic range of the radar waveform generator, transmitting, and receiving systems. High sensitivity has the side effect of revealing the echoes from many natural scatterers in the ionosphere, as well as the spectrum of ground clutter that has been spread in doppler by the fluctuations of the signal propagation path. In order to recognize and suppress these unwanted returns, which can obscure target echoes, an understanding of the underlying physics is essential.

A distinctive requirement of a skywave radar is a suite of auxiliary systems to monitor the state of the ionosphere and the availability of unoccupied channels in which to operate. This reliance on continuous environmental monitoring and the associated ability to adapt the radar parameters and tasking to make best use of the prevailing conditions cannot be satisfied with simple, low dynamic range equipment, for that would fail to reveal many of the phenomena that are setting the threshold for target detection and tracking. Further, a high degree of automation is essential to keep up with the changing environment.

A comparison of principle design parameters of several current and former operational HF skywave radars is presented in Table 20.2, illustrating the diversity of engineering solutions that have been implemented to meet specific mission objectives.

TABLE 20.2 Principal Design Parameters for Some Major HF Skywave Radar Systems, Past and Present

(This information has been compiled from sources that, in some cases, are incomplete. Where only partial information is provided, it may still be useful for comparison purposes. Notes: (i) VLPA denotes vertical log-periodic antenna and (ii) 2-band (*n*-band) linear arrays are usually constructed as contiguous collinear arrays.)

	Jindalee Stage B	JORN-Laverton	AN/FPS-118 ECRS/WCRS	AN/TPS-71 ROTHR	Nostradamus	Steel Yard (Komsomolskna Amur)
Developer	DSTO, Australia	Telecom Australia GEC-Marconi UK RLM USA-Australia	General Electric USA	Raytheon USA	ONERA, France	NIIDAR, Russia
Year of first target detections	1982	2000	1983	1987	1994	1977
Configuration	Quasi-monostatic	Quasi-monostatic	Quasi-monostatic	Quasi-monostatic	True monostatic	Quasi-monostatic
Tx – Rx site separation (km)	100	80	160	100, 160	n/a	
Max. transmitter power (kW)	160	300	1200	200	50	1500
Tx gain (dB)		18–27	23			
ERP	76 dBW	80 dBW	80 dBW	75 dBW		

TABLE 20.2 Principal Design Parameters for Some Major HF Skywave Radar Systems, Past and Present
(This information has been compiled from sources that, in some cases, are incomplete. Where only partial information is provided, it may still be useful for comparison purposes. Notes: (i) VLPA denotes vertical log-periodic antenna and (ii) 2-band (*n*-band) linear arrays are usually constructed as contiguous collinear arrays.) (*Continued*)

	Jindalee Stage B	JORN-Laverton	AN/FPS-118 ECRS/WCRS	AN/TPS-71 ROTHR	Nostradamus	Steel Yard (Komsomolskna Amur)
Tx array design	2-band linear array of 8 (low band) and 16 (high band) VLPA antennas	2 adjacent 2-band linear arrays of VLPA antennas; arrays oriented at 90° to each other	3 contiguous 6-band linear arrays of canted dipoles with 41-m tall backscreens, 12 elements per band; arrays oriented at 60°	2-band linear array of 2 × 16 VLPA antennas	Y-shaped array of 3 × 32 biconical elements randomly distributed	2 × vertical curtain arrays of 13 masts × 10 vertically stacked horizontal cage dipoles
Frequency range (MHz)	5–28	5–32	5–28	5–12 10.5–28	6–28	4–30
Tx apertures (m)	137	160, 160	304, 224, 167, 123, 92, 68		3 arms × 128 (long) × 80 (wide)	
Tx azimuth beam steer (deg)	± 45°	± 45°	± 30°	± 32°	360°	
Rx array design	Linear array of 462 5-m fan monopole pairs grouped as 32 overlapped subarrays	2 linear arrays of 480 twin monopole elements; arrays at 90°	3 contiguous linear arrays of 246 × 5.4 m vertical monopoles with 20 m backscreens; arrays at 60°	Linear array of 372 × 5.8 m vertical twin monopoles	Y-shaped array of 3 × 96 7 m biconical elements grouped as 3 × 16 subarrays with random distribution	2 × vertical curtain arrays of 30 masts × 10 vertically stacked horizontal cage dipoles
Rx aperture (m)	2766	2970	1518 / 2440	2580	3 arms × 384 (long) × 80 (wide)	500 (long) × 143 (high)
No. of receive channels	32	480	82	372	48	5
Rx azimuth beam steer	± 45°	± 45°	± 30°	± 45°	360°	
Waveform type	Linear FM/CW	Linear FM/CW	Linear FM/CW	Linear FM/CW Linear FM/CW	Coded pulse	Binary phase coded pulse
Waveform repetition frequency (Hz)	~ 4–80	~ 4–80	10–60	5–60		10, 16, 20
Waveform bandwidth(kHz)	4–40 typical	4–40 typical	5–40	4.17–100		40
Coherent integration time (s)	1.5– 5 Air mode 15–40 Ship mode	1.5–15 Air mode 15–40 Ship mode	0.7–20.5	1.3–49.2		
Primary mission	Aircraft detection	Aircraft detection	Aircraft detection	Aircraft detection	Aircraft detection	Ballistic missile detection
Secondary missions	Ship detection Remote sensing	Ship detection Remote sensing	Cruise missile detection	Ship detection	Ship detection Remote sensing	Aircraft detection Cruise missile detection

The term quasi-monostatic is used to denote configurations in which the transmit and receive sites are separated, as is often done with radars employing FM-CW waveforms, but not so far apart that the angle subtended at the target is more than ~5°, so the scattering behavior is close to what is observed for exactly monostatic geometry.

20.4 THE IONOSPHERE AND RADIOWAVE PROPAGATION

The solar activity that drives the ionization of the Earth's upper atmosphere is variable on a diurnal, seasonal, and long-term basis with a superimposed random component and occasional major storms and other disturbances. Further, the Earth's lower atmosphere is coupled to the ionosphere by a variety of upward-propagating wave and radiation processes, whereas the Earth's magnetosphere, the region beyond the ionosphere where the solar wind interacts with the Earth's magnetic field, is the source or conduit for corresponding perturbations from above. The ionospheric response to all these external forces is governed not only by inertial effects but also by chemical reactions and by the embedded time-varying electric and magnetic fields that link the ionospheric plasma to the Earth and to the interplanetary medium. As a consequence, the structure of the ionosphere undergoes changes on a wide variety of spatial and temporal scales, which drastically affect its properties as a medium for radiowave propagation.

The primary requirement for radar system design is a quantitative description of the propagation characteristics over the proposed coverage region. Specifically, the radar designer needs a statistical description that will enable matching the transmitted signal, power level, and antenna gain pattern to the supported frequency span, noise levels, propagation loss characteristics, and ray paths to the target region. In addition, the radar operator needs a model with enough sophistication to permit full interpretation of the real-time soundings for operating parameter selection, signal processing, and data analysis. With this latter requirement, a statistical description is usually not adequate, as important features may be lost. For example, under rapidly changing ionospheric conditions, the radar echoes will experience a time-varying doppler shift. Averaging over time would result in a doppler shift tending to zero. Clearly, this is of no value to the operator wishing to compensate for ionospheric motion and hence to retrieve a meaningful estimate of target radial velocity. As a second example, consider the situation where an atmospheric gravity wave (AGW) is propagating through the ionosphere in the vicinity of the control point (ionospheric reflection point) while a target is being tracked. The ionosphere is only ~ 0.1% ionized at the altitudes of interest to HF radar, but waves in the neutral gas, under the restoring force of gravity, transfer their motion to the free electrons via collisions. As the distribution of electrons defines the "reflecting surface" experienced by the radar signals, the apparent bearing and range of the target will fluctuate as the ionospheric "reflecting surface" undulates in response to the AGW. This undulation is known as a traveling ionospheric disturbance (TID). TIDs may have wavelengths of hundreds of kilometers and speeds up to 1000 kilometers per hour. Unless the radar makes appropriate real-time corrections to the target coordinates, tracking accuracy will be severely impaired.

To address these various needs, it is advantageous to adopt corresponding descriptions or models, emphasizing different aspects of the ionosphere and its influence on radiowave propagation and, hence, HF radar performance. In many cases of practical interest, ionospheric models developed originally for HF communications purposes

can be adapted to the radar context, where the main difference is the greater sensitivity of radar observations to dynamical processes. This arises primarily because of the extremely high dynamic range required to accommodate and preserve target echoes in the presence of strong clutter and external noise.

Ionospheric Structure. The basic physics of ionization and recombination processes leads to a natural division of the ionosphere into a number of regions:

D Region. This region occupies the lowest altitudes considered. It extends from 50 to 90 km, with electron density increasing rapidly with altitude in the daytime. Its properties reflect the balance between free electron production by the incident solar radiation flux and free electron loss via various electron-ion and electron-neutral recombination processes. Accordingly, the maximum ionization in the D region occurs near the sub–solar point and will be greatest during periods of highest solar activity (sunspot maximum), though it does not achieve densities sufficient to reflect or even significantly refract HF radio waves. The key role of the D region in HF radiowave propagation is signal attenuation via electron-neutral collisions that are frequent at these moderate altitudes where the neutral species density is still relatively high. It is not represented explicitly in some ionospheric models where its effects are accounted for with an empirically derived path-loss calculation.

E Region. This ionization region extends between about 90 and 130 km in altitude with a maximum near 110 km when sunlit. In addition, there may be anomalous ionization referred to as sporadic E. This latter ionization layer is typically only a few kilometers thick and usually short-lived, often lasting less than an hour; it may be either smooth or patchy, is seasonally and diurnally variable, weakly correlated with solar activity, showing a tendency to favor low sunspot numbers, and has marked variation with latitude. From the propagation perspective, sporadic E holds a special place as the layer providing the most stable propagation over coherent integration times typical of skywave radar. Because it is only ~100 km above the Earth's surface, the maximum range that can be reached via one-hop E-layer propagation is only ~2000 km, though normally the layer is not totally reflecting, so some energy continues upward to the F-layer where it is reflected to reach the Earth's surface at much greater ranges.

F Region. This is the highest-altitude region of interest for skywave propagation, and it is also the region of greatest electron density. In the daylight hours, the F region sometimes manifests two component layers, especially in summer. The Fl region lies between 130 and 200 km and, like the E region, is directly dependent upon solar radiation; it reaches maximum intensity about 1 h after local noon. The F2 region is variable in both time and geographical location. The altitude of the F2 region peak lies typically between 250 and 350 km at middle latitudes. The F2-region ionization shows marked day-to-day variations and, in general, is not the regular sun follower that the E and F1 regions are.

The simple physical picture of the ionosphere as consisting of several more or less concentric layers must be modified substantially at both low and high latitudes to take into account the effects of the impressed electric and magnetic fields. To begin with, the displacement of the Earth's magnetic axis from its rotational axis means that the ionosphere does not preserve a more or less constant form enveloping a rotating earth. Near the magnetic equator, where the geomagnetic field is close to horizontal, atmospheric tides and associated winds drive the so-called E-region and F-region

dynamos, resulting in an upward drift of the ionospheric plasma and its subsequent descent along the geomagnetic field lines. The resulting electron density depletion at the equator and roughly symmetric enhancements in the vicinity of ~ 20° N/S latitude is known as the *Appleton* or *equatorial anomaly.* Besides causing errors in estimated target range and bearing, the tilted ionospheric "reflecting surface" near the equator can support scattering of incident radiowaves into low-loss elevated trans-equatorial modes (chordal modes) that return strong clutter from the opposite hemisphere, often with highly unstable phase characteristics.

In the polar zones, the near-vertical geomagnetic field lines provide a pathway for charged particles and disturbances of solar and magnetospheric origin to reach ionospheric heights and contribute to ionization processes and plasma transport. The best known phenomena here are the aurorae, which are concentrated in ovals poleward of the boundary where the Earth's geomagnetic field lines change from closed (connected to their images in the opposite hemisphere) to open, that is, connected to the interplanetary magnetic field. A wide variety of plasma waves and instabilities populate these regions, producing irregularities that achieve high electron densities and hence, provide strong sources of spread-doppler clutter. They have been known to impact skywave radar systems severely.

Ionospheric Variability. While all ionospheric properties are time-varying, from an OTHR perspective, it is useful to separate the "fast" processes or "dynamics" from the "slow" or "structural" variability. A phenomenon is termed *dynamical* relative to a radar observation process if it occurs on a timescale commensurate with the corresponding process timescale, such as (i) pulse/waveform repetition interval, (ii) coherent integration (dwell) time, (iii) scan revisit time, or (iv) mission / track lifetime. Dynamical processes impact directly on how one should process the received signals or perform tracking. Slow processes, such as the 11-year solar cycle, seasonal changes, and the diurnal cycle of the E- and F-layers can generally be treated as quasi-stationary background processes that set up the ionospheric structure at any given time, within which fast processes may occur. An exception to this classification arises with the dawn and dusk terminators, that is, the day-night boundaries; sweeping around the Earth at 1600 km/hr, they produce abrupt changes in the ionosphere and trigger large-scale instabilities.

Structural Variability. The day-night cycle produces drastic changes in the ionization distribution within the ionosphere. At night, the D-layer disappears, the E and F regions experience a substantial decrease in ionization, and the equatorial and polar regions are more prone to large-scale perturbations. The extent of diurnal variation can be seen by examining Figure 20.4, which shows measured electron density (expressed in terms of plasma frequency as defined in Eq. 20.1) versus virtual height and time-of-day at a mid-latitude location.* Typically, the diurnal variation requires a radar to vary its operating frequency by more than an octave in frequency to maintain 24-hour surveillance over a fixed target location.

---

* Virtual height is the reflection height computed from signal time delay by assuming that the radiowave travels at the speed of light as if in free space; in fact, the radiowave group velocity in the plasma is lower, so the true height is less. For most purposes, it is more convenient to deal with virtual height, because there are two theorems that dramatically simplify practical calculations. Martyn's theorem shows that complicated oblique skywave ray paths via the true reflection point can be replaced to good approximation with simple rectilinear geometry via the virtual reflection point. The theorem of Breit and Tuve demonstrates that, to good approximation, the time-of-flight is unchanged by this substitution. Davies[21] provides a clear explanation of these useful theorems.

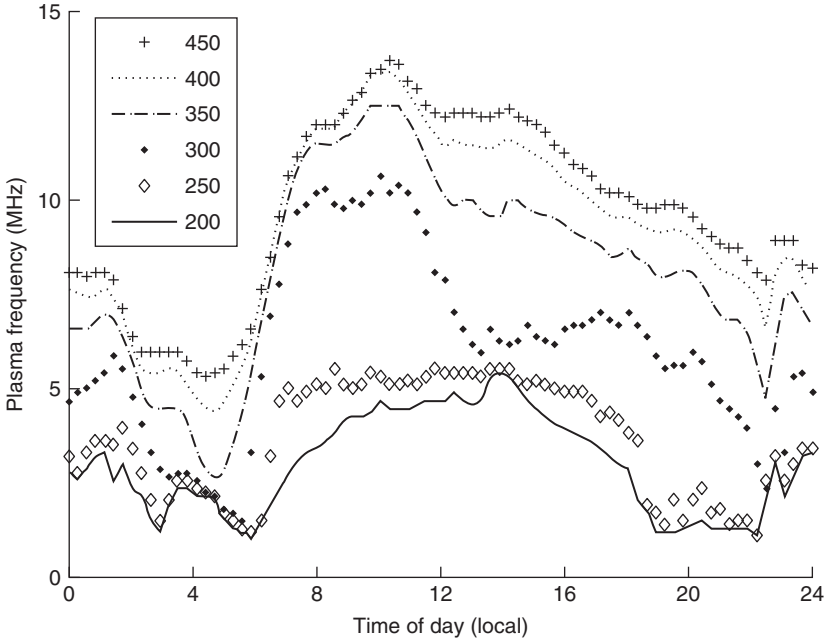


FIGURE 20.4 Diurnal variation of the electron density profile, as measured by the plasma frequency, plotted for various virtual heights (in km). The data was recorded by a vertical incidence sounder at latitude 18.0 S, longitude 144.9 E on 17 September 2002, SSN = 88.

What is more surprising is the day-to-day variability, even at mid-latitudes, which impacts significantly on HF radar performance. The inability to predict reliably even a day in advance is a serious consideration in radar design and scheduling.

An illustration of day-to-day variability is presented in Figure 20.5, which overlays 30 vertical incidence soundings recorded at the same time of day for a month.

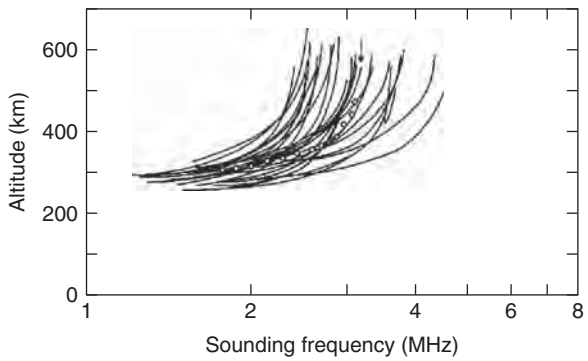


FIGURE 20.5 Comparison of measured vertical incidence ionograms for a fixed time of day over a month with a model-based prediction of the median, denoted by the small circles.

The traces plot the virtual height versus radiowave frequency for the *ordinary* ray (see subsection below on radiowave propagation). Soundings of this type—ionograms—measure the return trip time delay for a signal to travel up to the height at which the electron density is sufficient to reflect it, that is, where the plasma frequency f_p equals the incident radiowave frequency.

Also shown in Figure 20.5 is the corresponding monthly median, taken from the model of Thomason et al.[22] Consider the critical frequency, that is, the highest reflected frequency, corresponding to the ionosphere's peak electron density. The upper and lower deciles depart from the median by typically ± 25%. Over such a range of frequency, terms in the radar equation such as antenna gains, target RCS, and slant range may vary substantially, so radar performance is inevitably statistically distributed. Although it makes sense to use median values for many radar performance modeling calculations, the radar designer should adopt a conservative approach and assume the lowest critical frequency.

Perhaps the most intriguing major cause of systematic variation is the 11-year cycle of solar activity. Various correlated parameters have been defined to measure this activity, including the 10.7 cm solar flux, sunspot number, and various magnetic indices. Figure 20.6 shows the smoothed average sunspot number plotted from 1956, the year of the first validated OTHR detections of military targets (achieved by the *MUSIC* radar at NRL[2]).

Enhanced solar activity impacts on the ionosphere in many ways, but its most significant action from a radar perspective is to produce substantially higher ionization levels that persist at usable levels through the night, so higher radar frequencies can be employed and the minimum achievable range decreases. In addition, the height of the maximum electron density increases, so one-hop propagation can reach greater ranges.

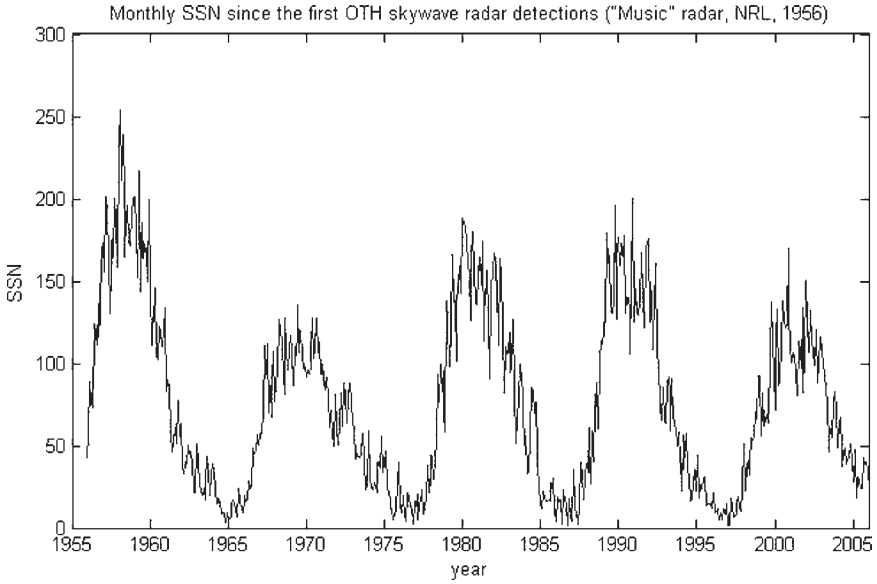


FIGURE 20.6 The variation of the monthly median sunspot number since 1956, the year of the first operational skywave radar detections

Daytime absorption increases for a given frequency because of higher D-layer densities, but this is compensated by access to higher operating frequencies. An increase in the number and severity of magnetic storms and sudden ionospheric disturbances is another consequence of higher solar activity, as are increased post-sunset plasma bubble activity and associated scintillation in equatorial regions, solar flares, and coronal mass ejections all of which disrupt stable propagation. Solar flares are giant explosions on the sun's surface, generally occurring near sunspots, which emit ionizing radiation that penetrates into the D region and dramatically increases absorption. The resulting *sudden ionospheric disturbance* (SID) or *short-wave fade-out* can completely incapacitate a skywave radar for a period from minutes to hours. Perhaps more importantly, although the flare itself is not predictable, the associated burst of particles will start to arrive a few hours later, with most ejected material reaching Earth several days later, seriously disrupting HF propagation. The critical importance of solar activity will be demonstrated in Section 20.13 in the context of radar performance prediction.

Ionospheric Dynamics. As the radar signals traverse the ionosphere, motions of the plasma medium along the propagation path imprint themselves on the signals in ways that can degrade or obliterate the target information of interest. Skywave radars can be designed to recognize the signatures of these phenomena, adjusting the radar frequency, choice of waveform, and processing to mitigate their effects where necessary. Sometimes the ionospheric motions actually help the radar, for example, by enabling it to discriminate between natural noise and some forms of deliberate interference. Thus, it is vitally important to understand not only the structure of the ionosphere and the opportunities it provides for propagation, but also the motions and disturbances that inhabit it.

The variety of wave types and irregularity of production mechanisms in the ionosphere is enormous, and includes not only those arising naturally but also many induced phenomena such as those resulting from ground-based ionospheric RF heaters or rocket interactions with the ambient plasma. Among the most important from the radar point of view are the following:

- Transient plasma structures associated with ionized meteor trails; at any given location, meteors have a strong diurnal variation and directionality. They are ever-present and constitute a major source of clutter for HF radars.[19]
- Large and medium scale atmospheric gravity waves (AGW) produced by energetic phenomena in the lower atmosphere propagate upward with increasing amplitude until nonlinear processes begin to dominate, resulting in wave breaking. Large AGWs may persist for hours and propagate over global distances, causing serious deviation of incident radiowaves.[23] They are sometimes the major cause of tracking errors.
- Magnetic disturbances originating where the solar wind impacts on the magnetosphere; they propagate earthward and cause resonant oscillations of the geomagnetic field lines permeating the ionosphere. The field lines are "frozen in" to the ionospheric plasma and, hence, the ionosphere vibrates locally at frequencies typically in the range of 10^{-2}–10^{0} Hz, imposing a corresponding modulation on any transiting radiowaves.[24] Ship echoes can be obscured by this modulation.
- The equatorial electrojet, part of the global system of fields and currents driven by the dynamo action of winds and tides, is host to small scale field-aligned irregularities with characteristic velocity distributions.[25] Some small-scale plasma instabilities may endure for a fraction of a second, others for tens of seconds, evident to radars as spread-doppler clutter that prevents the detection of small targets.

- Equatorial plasma bubbles often appear after sunset and convect upward through the F region where they contribute to strong diffuse doppler-spread scattering referred to as spread-F.[26] Detection of all targets is difficult when this effect occurs.
- The aurorae are dynamic structures governed by fields and plasma flows in the magnetosphere; strong electric fields aligned with the magnetic field accelerate electrons down into the ionosphere where they produce highly ionized formations that reflect radio waves very efficiently. The resulting doppler-spread echoes are so strong they can incapacitate a radar through its sidelobes.
- Geomagnetic storms and sub-storms result from solar flares and coronal mass ejections. As mentioned earlier, hard X-rays and ultraviolet radiation raise D-layer ionization dramatically with consequent increases in radiowave absorption. Within an hour and for a day or so after, flare-generated particle bursts start to arrive and are channeled down the magnetic field lines at high latitudes, causing ionospheric heating and associated diffusion to lower latitudes, together with a variety of magnetic field perturbations. HF propagation is often severely disrupted.

Models and Their Uses. Knowledge of the conditions to be expected for a particular radar deployment is vital to radar design, as well as providing a guide to echo interpretation and a means to simulation and performance prediction. This type of information, based on decades of ionospheric observations and theory, is conveniently distilled in models that are widely available and used extensively. But even more importantly, from the operational perspective, HF radars must maintain a real-time ionospheric model (RTIM) that is intimately linked with the radar subsystems, serving to guide frequency selection, radiated power, task scheduling, coordinate registration (converting from the radar coordinates of time delay and angle-of-arrival to geographical coordinates), ionospheric mode structure interpretation, and association of multiple tracks from a single target. Unlike the climatological models, RTIMs must be updated continuously with information from an auxiliary network of beacons. Oblique and vertical incidence sounders and transponders have been deployed for this purpose. For instance, the JORN radars exploit approximately 20 sounders and related facilities distributed around the coast of Australia. Radar performance in most operational roles is governed by the fidelity of the adopted RTIM.

It is important to differentiate between models that describe the physical (or physico-chemical) state of the ionosphere and models that describe radiowave propagation characteristics, though the latter are often derived from the former by applying ray-tracing methods, and the former are predominantly derived from radiowave propagation measurements such as point-to-point link statistics and vertical incidence soundings. Both classes are of general applicability to HF communications and geophysical investigations, as well as skywave radar.

Models of the Ionospheric Medium. Models of the ionosphere fall into two categories:

- Climatological models based on sounder, rocket, and satellite measurements. Being derived from statistics, they provide no explicit information on real-time "'weather." that is, irregularities, waves, and other dynamic processes, though measures of variability may be provided. Many early models had their genesis in the large database of recorded ionospheric soundings made during the International Geophysical Year of 1957–1958 and the International Year of the Quiet Sun of 1964–1965. These models,

which focused on the spatial distribution of electron density, were used extensively in early HF radar performance analyses and include ITSA-1, ITS-78, RADAR C, IONCAP, and AMBCOM.[27–32] Lucas [33] provides details of these models and their origins. Some of the prediction methods have not been well documented although widely distributed; also, users frequently "improve" upon a model and prediction method to suit their specific needs. As an example, the model RADAR C[29] is the basic building block of Thomason et al. in NRL Report 8321[22]; however, they added a D region, a collision-frequency distribution, an Earth's magnetic field, a topside electron distribution, an auroral electron-density modification,[34] and other features that make the model more generally useful. The ionospheric model as described in NRL Report 8321 has been used for some of the examples presented in Section 20.13.

The International Reference Ionosphere (IRI) is perhaps the foremost modern example, presently available as version IRI-2001.[35–37] Other climatological models used in HF radar applications are PIM,[38,39] PRISM,[40,41] and FAIM.[42]

- Physics-based or first principles models, such as USU GAIM,[43] JPL/USC GAIM,[44] SAMI3 developed by Huba and Joyce,[45] and Khattatov's model,[46] solve the plasma dynamics and composition equations governing evolution of density, velocity, and temperature for various ion species on a global 3D grid, subject to the Earth's magnetic field and prevailing solar indices. These models required seeding with initial conditions, often a limiting consideration.

For both classes, improved accuracy of forecasting is achieved by assimilating data from ground-based sounders, total electron content (TEC) derived from GPS, UV airglow data, and in situ measurements of electron density from satellites and other sources. Often the assimilation is performed within an extended Kalman filter framework, so accuracy estimates are a byproduct. Further, site-specific applications may benefit from adaptation of the underlying model parameters and coefficients.

The complex structure and dynamics of the ionospheric medium govern HF skywave propagation primarily through the space-time variation of the free electron density distribution. A useful simplification is to regard the large-scale ionospheric structure as defining the propagation geometry throughout the illumination volume, whereas the dynamical processes impose their respective modulations on the transiting signals.

Computational Aspects and Ray-tracing. For many OTHR purposes, it suffices to employ a ray-theoretic representation of the radiowave field. Ray-tracing techniques fall into two categories: analytic and numerical. Analytic methods are fast but rely on fitting parametric models to the electron density profiles and are, hence, of limited use for operational applications where accuracy is critical. They are also limited by their inability to handle magnetoionic effects on propagation. Nevertheless, they provide closed form expressions for group range, phase path, ground range, and other parameters, and therefore, can be very useful for computationally intensive studies such as system optimization. The multi-quasi-parabolic (MQP) model of Hill[47] based on Croft's QP technique[48] is widely used, while quasi-cubic models have been proposed by Newton et al.[49]

Numerical ray-tracing codes are versatile and able to accommodate almost arbitrary ionospheric structure, at the expense of computational burden. Of the many numerical ray-tracing codes developed over the years, the implementation by Jones and Stephenson[50] based on integration of the first-order Haselgrove equations remains the most widely used. Coleman[51] has developed an alternative implementation.

Models of Ionospheric Radiowave Propagation. From the HF skywave radar perspective, interest is usually centered not on the ionosphere per se but on how it determines radiowave propagation. Models such as VOACAP[52] and ASAPS[53] generate predictions of point-to-point (circuit) parameters, including maximum and minimum usable frequency, elevation angle, group path, mode probability, path loss, and signal-to-noise ratio, for user-specified terminal characteristics. A more general capability is available with Proplab.[54]

As these models are based on their respective climatological databases, they are of no use for real-time applications such as RTIM and are inadequate for serious radar design, which should be based on measurements taken at the proposed deployment site, but they can be useful for answering certain questions, such as "Can my signal be heard at X and if so, at what SNR?"

For calculations requiring moderate accuracy, geometrical optics or virtual ray-tracing based on Martyn's Theorem can be applied to stored semi-empirical electron density profiles, or better, stored "snapshots" of the ionosphere generated by an RTIM. Alternatively, analytic ray-tracing can be performed on analytic profiles fitted to RTIM databases. The most accurate predictions come from the application of sophisticated ray-tracing routines to a database of RTIM snapshots. When a radiowave propagation model is combined with radar system parameters, target scattering characteristics, and HF noise distributions, the radar equation (Eq. 20.2) can be solved to predict the radar's performance, as treated in Section 20.13. This is the basic approach employed within RADAR C,[29,55,56] for instance. From the operational HF radar viewpoint, use of these propagation models is limited to statistical studies of radar performance, not real-time radar support applications such as coordinate registration because of the poor fidelity of the assumed electron density distributions and the shortcomings of geometric ray-tracing.

Other Models and Propagation Issues. Studies of HF radar performance at low latitudes have shown that it is often necessary to incorporate models of dynamical processes, either because they manifest themselves directly in the doppler structure of radar echoes or because they are indicators of other phenomena that do. Useful models in this category include HWM93,[57] which describes the structure of the zonal and meridional neutral winds throughout the ionosphere, and WBMOD,[58] which describes scintillation arising from small scale irregularities such as those associated with spread-F in the post-sunset ionosphere. These models find important applications to the analysis and interpretation of doppler-spread clutter.

A number of phenomena have been ignored in the preceding discussion, though their effects can be observed in some skywave radar systems. They include (i) a variety of nonlinear processes that can occur during ionospheric propagation,[59] (ii) delayed echoes,[60] focusing at the antipode, and (iii) round-the-world propagation.[61] The most significant practical exploitation of such phenomena can be found in experiments directed at ionospheric modification.

20.5 WAVEFORMS FOR HF RADAR

The factors that govern the choice of waveform in HF radar systems can be grouped into two classes. First, there are the considerations common to microwave radar, that is, range and doppler resolution as described by the ambiguity function and optimized

for target detection and estimation, realizability in hardware, susceptibility to interference, efficiency, and the electrical properties of the scatterers of interest. In addition:

- HF radar waveforms must fit within available clear channels in the HF frequency band, complying with stringent constraints on leakage into channels occupied by other users.
- They must be compatible with the fact that the radar is operating in a "waveguide," namely the volume between the Earth's surface and the ionosphere, with unique possibilities for multiple propagation paths including round-the-world propagation, significantly modifying the effective ambiguity function.
- They must be able to achieve the desired measurement capability in the presence of extremely strong ground clutter.
- They must be designed so as to minimize distortion or corruption by the ionospheric medium or, at least, enable such distortion to be estimated and mitigated by signal processing after reception.
- They must heed the constraints on peak versus average power imposed by the HF transmitting equipment and antennas.

The waveforms used in most operational HF skywave radars are variations on the periodic linear frequency-modulated continuous wave (LFM-CW) signal. Often, there is some provision for amplitude shaping, normally at the commencement and end of each sweep. The Jindalee radar was designed with the facility to apply a number of amplitude notches within the sweep, thereby enabling the radar to sweep at zero amplitude across narrow-band users in the same frequency band without causing interference. Another class of variations involves departing from a linear frequency modulation. By varying the frequency-time characteristic of the waveform, range sidelobes can be reduced and spectral leakage can be controlled. Controlling the phase discontinuity from the end of one sweep to the beginning of the next provides another dimension in which the waveform properties can be optimized. Further generalization of the FM-CW waveform is possible by relaxing the condition that the waveform be periodic. This is a powerful tool for controlling range-ambiguous echoes, which can be shifted about in the range-doppler plane to uncover previously obscured target echoes. And perhaps most importantly, in the congested HF spectrum where clear channels of adequate bandwidth to achieve the desired resolution may be scarce, FM-CW waveforms defined over two or more separate subbands are readily synthesized.

Most early HF skywave radars employed pulse waveforms, in part because the technology of the day did not support FM-CW waveforms with the required level of spectral purity for this demanding application but also because pulse waveforms enjoy some undeniable advantages. First, they can be employed from a single transmit-receive site, avoiding the cost and complexity of acquiring suitable land, duplicating many facilities and synchronizing two widely separated sites. Second, the ability to gate the echoes in time means that the only clutter power that impacts on the range footprint spanned by a pulse is clutter originating in that footprint. This relaxes the waveform generator dynamic range requirements somewhat, especially when phenomena such as auroral clutter may cause problems, though to achieve the same probability of detection, peak power must increase to maintain equivalent average power. Third, as a consequence of reduced dynamic range requirements and assuming for the moment that strict spectral emission controls are not an issue, more efficient amplifiers can be used, as discussed in the following section. Fourth, pulse waveforms may be less susceptible to some forms of jamming.

But there are disadvantages. First, emission controls are almost invariably a serious matter when it comes to obtaining a license to radiate. Second, there are radar applications beyond basic target detection where extremely high spectral purity is essential, as with detection of small ships at low speeds. Third, the antenna design must be able to handle higher field strengths without arcing and sparking, which introduces noise. Fourth, unlike the FM-CW case, it is not generally feasible to synthesize suitable pulse waveforms from separate subbands of the HF spectrum when wide clear channels are not available. Fifth, there are fewer options for radiating multiple waveforms simultaneously from the same transmitting facility. And sixth, for high power HF radars, the additional power density associated with pulse waveforms in the ionosphere may, in principle, cause self-modulation from nonlinear effects.

20.6 THE TRANSMITTING SYSTEM

Transmitters. Most of the radar designs and missions require transmitter average power levels between 10 kW and 1 MW. Antennas are generally arrays of radiating elements, and the common practice with HF radar is to drive each element with a separate amplifier. This approach permits beam steering at a low power level in the amplifier chain. The active element in each final transmitter stage can be either a traditional vacuum tube[62] or a solid-state device.[63,64] Most operational HF radars employ solid-state amplifiers based on a hierarchy of modules, starting with elemental amplifiers of perhaps 500 W and combining these progressively via passive networks until the final output power is attained. Relative phase shifts or time delays are inserted in the amplifier chain, driving each antenna element to steer the resultant beam. This architecture enhances reliability and provides graceful degradation in the event of module failure.

Solid-state HF radar transmitters operate at power efficiencies lower than those based on vacuum tube amplifiers. Vacuum tube amplifiers are also more robust and have been used successfully in a number of HF skywave radars including the AN/FPS-118. The adoption of solid-state amplifiers in radars such as ROTHR and Jindalee/JORN is driven by the need to accommodate instantaneous frequency switching over wide bandwidths while maintaining high linearity and spectral purity. These radars routinely interleave different surveillance tasks with widely separated carrier frequencies—8 and 24 MHz, for example—switching as often as every one or two seconds while conforming to strict spectral emission standards. Such instantaneous frequency changes would place unachievable switching demands on the high-level vacuum tube radio frequency circuits. Meeting power control and amplitude-shaping of waveform requirements dictates linear operation of amplifiers.

Apart from the requirement to meet emission guidelines laid down by national and international spectrum management authorities, high spectral purity is essential because skywave radar uses doppler processing to separate the targets from the clutter, and hence, the clutter returned on the phase and amplitude noise sidebands radiated by the transmitter must be kept below the echo power of desired targets. This can impose a stringent condition on the emitted signal-to-noise ratio of the transmitter, and hence, on the signal-to-noise ratio of the waveform generator. For example, noise spectral density at 10 Hz from a carrier may need to be as low as –100 dBc in order to detect some targets of interest. For designs employing an amplifier for each

antenna element, the radiated phase noise will generally add noncoherently and thus be suppressed by ~15 dB relative to beamformed noise power for a 32-element array. But that is not the end of the matter—transmitter phase noise will be reflected by the distant Earth surface and integrated across the receiver passband, raising the phase noise contributions to the noise floor by a factor that, for an FM-CW waveform, is roughly equal to the ratio of the waveform bandwidth to the waveform repetition frequency. The lower-level stages of signal amplification can generally be designed to add essentially zero noise, but mechanical vibration in the high-power amplifiers can add appreciable amounts so care must be exercised in the air or liquid coolant flow system design.

If the radar is to perform wide-area surveillance, frequent frequency changes are required in order to cover the various range extents. In addition, relative phase or time-delay changes are required in each amplifier chain to accomplish azimuthal steering. A broad-bandwidth performance and a tolerance to a variable voltage standing-wave ratio load are, therefore, essential features in an HF radar transmitter. Typically, one might specify VSWR < 2 at full power over an operating range of 5–30 MHz, say, with an ability to tolerate higher VSWRs at fractional power outputs. Since the antenna elements will be wideband, harmonic filters may be required. For example, one transmitter and harmonic filter combination might have a 5 to 9 MHz passband and a stopband for 10 MHz and higher frequencies; a second combination might pass up to 17 MHz and reject 18 MHz and higher, and the design would continue in this manner to the highest frequency of operation. Some skywave radar designs call for up to six bands. A related issue is the occurrence of mutual coupling between antenna elements in the transmit array.[65] Energy coupled back into an amplifier from its neighbors can reach levels that result in intermodulation distortion, as well as causing load resonances that stress the amplifier chain.

Antennas. The choice of antenna configuration is intimately linked to the radar mission, generally defined in terms of target types, radar coverage, and coverage rate. The Naval Research Laboratory magnetic-drum recording equipment (MADRE) radar[2] employed a single antenna, duplexed and used for both transmit and receive. This 100-m-wide by 40-m-high aperture provided sufficient gain and angular resolution for aircraft tracking in the upper part of the HF band. The experimental French OTH radar Nostradamus[11–13] likewise employs a single antenna array, configured as three horizontal arms of length 384 m radiating from a central control center, though only a subset of elements are used to transmit, whereas all of the elements are used for reception. Such truly monostatic designs have the special advantage that the outbound propagation path to a target is almost identical to the in-bound path. Monostatic radars with separate transmit and receive antennas experience decorrelation between the paths as the separation increases and different parts of the ionosphere become involved. Either form of monostatic radar avoids the costs of multiple sites and associated communications infrastructure, and the challenge of finding suitable sites with an appropriate geographical relationship, but they are constrained in waveform choice and/or radiated power by the need to avoid simultaneous transmission and reception, as well as being potentially susceptible to range-folded clutter. In particular, the class of FM-CW waveforms has been widely adopted, driven mainly by constraints on spectral emissions outside the nominal radar bandwidth. For these reasons, configurations employing separate transmit and receive antenna sites are employed in radars such as Jindalee, JORN, and ROTHR, usually in a quasi-monostatic arrangement where the

intersite separation is about 50–100 nautical miles, much less than the range to the target zone but sufficient to prevent self-jamming when using continuous waveforms and to separate the transmit and receive range ambiguity zones in azimuth.

Transmit radiating element choice is driven primarily by the range of frequencies to be radiated and the waveform bandwidth, but it must also take into account the range and azimuth coverage required, the associated coverage rate, and concerns about clutter, especially spread-doppler clutter. The vertical radiation pattern controls these issues. The power-handling capabilities of the antenna elements are also a consideration.

A lower bound on the transmit array aperture is set by the need to achieve adequate directivity and hence power density on the target; the sensitivity required depends on the size of the targets of interest. The upper bound is often set by the revisit requirement—in general, the radar will step over a wide arc, but it must sample each region frequently to maintain tracks on maneuvering targets, so the transmit beamwidth should not be too narrow. The need to keep VSWR to modest levels is usually addressed by having 2 to 6 arrays addressing subbands of about one octave of frequency each.

Some of these considerations are influenced by the advantages that attach to employing multiple simultaneous receive beams, so a broad but well-shaped transmit beam that can be filled by perhaps 10 to 30 narrow receive beams is a popular choice. In the limiting case, the transmit array may floodlight the entire sector of coverage, which is filled with receive beams that stare continuously. The nonuniformity of clutter sources can cause serious problems with any design that does not have control over the array pattern. Further, HF antennas seldom achieve high front-to-back ratios, so the risk of signal masking by backlobe clutter cannot be ignored. For linear arrays, offsetting the transmit and receive array boresights is a moderately effective measure that must be traded off against the concomitant reduction in main beam overlap. Linear arrays of vertically polarized log-periodic antennas, vertical planar arrays of horizontal dipoles, stacked Yagi antennas, elevated rhombic elements, linear arrays of tilted monopoles, and a two-dimensional array of biconical antennas have all been used in skywave radar transmit systems, in some cases, with backscreens to improve the otherwise mediocre front-to-back ratios.

In the elevation plane, desirable radiation angles run between 0° and 40° for common values of range and reflection height. The vertical beamwidth needs to be sufficient to illuminate the required range depth; in general, this is automatically satisfied owing to the cost and complexity of an antenna able to form a beam narrower than this in the vertical plane. For most scenarios, any sensitivity gained by directivity in elevation directly improves radar performance, since instantaneous range depth is generally limited by ionospheric effects. This is in contrast with azimuth directivity for transmitting, where an increase in directive gain is accompanied by a decrease in area coverage. For noise-limited detection, this can be compensated by a reduction in dwell time, but for clutter-limited detection there may be a penalty for coverage rate. Some radars use horizontal two-dimensional arrays with up to ~ 100 receive channels to achieve quite high vertical directivity on both transmit and receive. In the past, others, such as MADRE and some former Soviet Union skywave radars, employed vertical two-dimensional arrays, up to 143 m high and 500 m wide in the case of the Soviet radars.

At low elevation angles, the antenna pattern is strongly influenced by the electrical and magnetic properties of the ground around and in front of the array. To achieve

gain at low-angles, it is standard practice to install a ground mesh screen; this has the secondary benefit of avoiding pattern distortion due to inhomogeneities in the soil. For example, the Jindalee transmit arrays sit on ~ 80 hectares (~ 200 acres) of steel mesh, extending ~ 200 m in front of the arrays.

Notwithstanding the merits of vertical directivity, most skywave radars do not employ steerable directivity in elevation but cover all necessary radiation angles with one broad elevation beam. This choice permits the antenna to have a relatively small vertical dimension, and hence reduced cost, though demands on antenna radiation efficiency impose a lower limit.

Another issue is the choice of transmit polarization. Nearly all operational skywave radars radiate with vertical polarization, based on the ease of achieving good vertical coverage at moderate cost by means of the widely used log-periodic broadband antenna. The large curtain arrays of the former Soviet radars used horizontal dipole elements whereas the USAF AN/APS-118 used inclined dipoles to adapt to ground conditions. The potential benefits of full polarization control on transmit have been assessed and a number of experimental studies carried out, but no operational system has gone down this path.

The antennas and power amplifiers used in HF broadcast stations have much in common with HF radar, that is, to maintain a specified level of illumination over a designated area. To achieve this goal, many multiband and steerable broadcast antennas[66–68] employ large vertical apertures. Antennas used for HF radar have an additional severe constraint: the need to minimize mechanical motion due to the wind (*Aeolian vibration*) that would cause signal phase modulation that would then be imposed on the transmitted signals; this requirement is easier to meet with low-antenna-height designs. An under-exploited advantage of broad beams in elevation is the ability to illuminate an extended range depth when conditions permit, providing clutter maps outside the range band of immediate interest for target detection; these can be used to schedule subsequent surveillance tasks.

20.7 RADAR CROSS SECTION

The radar scattering properties of targets determine both their detectability and the prospects for target classification; therefore, much effort has gone into establishing precisely how to describe those properties for a given radar context. In the HF skywave radar case, the inevitability of Faraday rotation that occurs during ionospheric propagation has been used to argue that a fully polarimetric treatment is unnecessary, so it is common practice to represent the scattering behavior in terms of scalar radar cross section (RCS). Fully polarimetric formulations are relevant when modeling complex scattering processes[69,70] and for target classification studies.[71]

In general, aircraft and ships have dimensions that put them within the resonant scattering regime, though the smallest aircraft and cruise missiles will lie in the Rayleigh scattering regime for the lower half of the HF band. Here, the RCS displays limited aspect sensitivity and a strong dependence on the target's gross dimensions. For an aircraft, the span of the wings, the fuselage length, the tail and elevator span, the vertical stabilizer and rudder height, and their relative locations are the main features that influence the RCS. Target shaping on a scale size much less than a wavelength will have little effect. Accurate measurements of radar scattering at HF frequencies is challenging, but facilities for making scale model measurements are

widely available, including anechoic chambers, indoor compact ranges, and outdoor test fields.[72] Full-scale experimental measurements can be made in some circumstances by means of calibrated reference scatterers or transponders deployed in the target zone and modulated to separate their returns from the target and clutter echoes. For bodies with highly conductive surfaces, the scattering cross section can be calculated quite accurately by numerical methods, such as the method-of-moments code NEC.[73] As a rule of thumb, the HF RCS of aircraft can usually be computed to an accuracy of about 2 or 3 dB (with respect to measured values) without resorting to highly sophisticated techniques.

When precise RCS information is not essential, rough but useful RCS estimates can be made by examining the scattering behavior of a few "canonical" shapes. Figure 20.7 is a family of plots giving RCS versus radar frequency for an oblong-shaped conducting body. The straight line marked 90° λ/2 dipole gives the RCS of a resonant, conducting half-wavelength rod, where the rod is parallel with the electric field. This geometry gives the maximum RCS for the rod. The upper scale of the abscissa gives the one-half-wavelength dimension of the frequency given on the lower scale. The curve marked 90° is the RCS of the oblong-shaped conducting body of 11 m length and 1 m thickness; again, the target long dimension is aligned with the electric field. The maximum RCS coincides with the nominal half-wavelength dimension or with the first resonance.

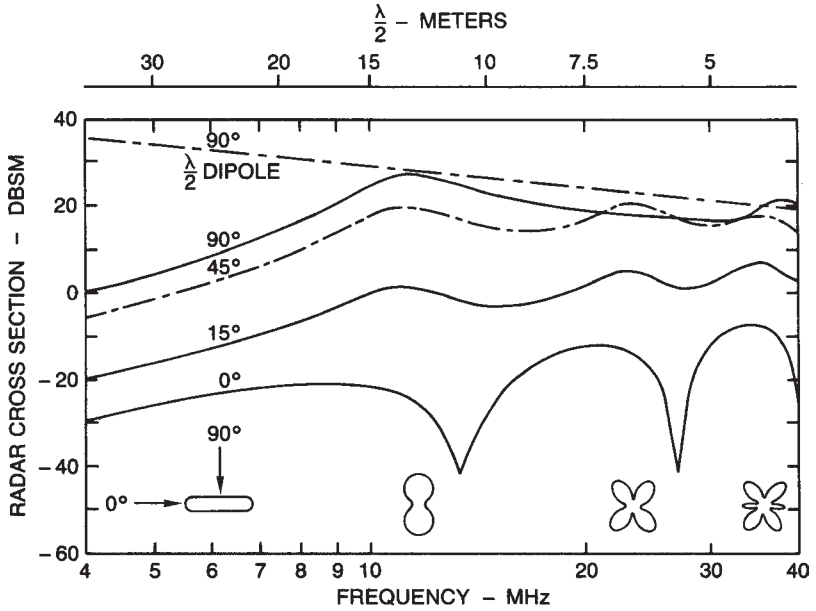


FIGURE 20.7 The frequency dependence of the RCS for an 11 m long by 1 m diameter, perfectly conducting cylinder, presented for various illumination geometries. The E vector and the 11 m dimension are in the same plane; 0° (nose-on), 45°, and 90° (broadside) curves are shown. The top (dashed) curve is for a resonant dipole at 90°. The small sketches at the first, second, and third resonances show the RCS angular pattern near these frequencies.

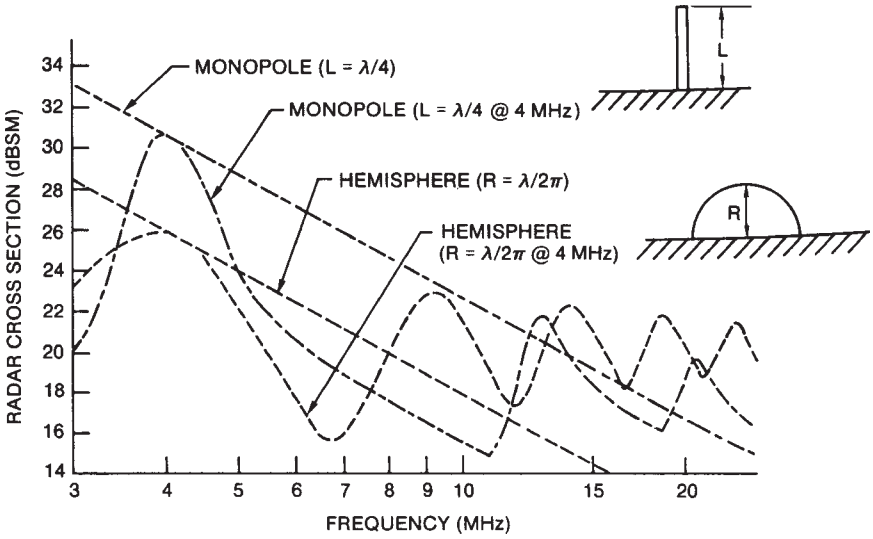


FIGURE 20.8 RCS frequency dependence of a rod (monopole) and a hemisphere on a perfectly conducting plane for vertical incident and scattered polarization

The curves marked 45, 15, and 0° give the RCS as the target is rotated to these angles in the plane that contains the electric vector. The little sketches give, at the left, the body shape, and then the RCS patterns at nominal 1/2 wavelength, 1 wavelength, and 3/2 wavelength, in order to help visualize how the RCS will change as the aspect angle is varied. For targets of other lengths with approximately the same shape factor, the response can be determined by sliding the curve along the λ/2 line and making the first resonance coincide with the line at the 1/2 wavelength point. As has been mentioned, Faraday rotation results in varying incident polarization, so over time a target will experience both favorable and mismatched polarization, with resultant fading of the scattered signal. Of course, additional fading of the scattered signal occurs due to time-varying polarization mismatch at the receiving antenna.

Figure 20.8 gives the vertical-polarization RCS of a rod and a hemisphere mounted on a perfectly conducting surface. With these canonical shapes, an estimate of RCS can be made for surface craft by matching L and R to the principle dimensions to the target. For small vessels, the mast height will be of most importance.[74] For surface targets, where the maximum RCS occurs with vertical polarization, a 12 dB skywave RCS enhancement results from the image field.

To illustrate the RCS behavior of a typical aircraft in more detail, Figure 20.9 shows the RCS of the F-18 fighter, as computed here by NEC2 applied to a wire-grid representation of the aircraft derived from a plastic kit model. The RCS shown is for monostatic (backscatter) geometry and horizontal copolar (HH) polarization. Calculations are presented for frequencies of 12, 18, and 30 MHz.

Although the fidelity of such RCS calculations of conventional platforms has been confirmed on many occasions, it is not clear that standard computational methods are applicable to the problem of estimating the RCS of aircraft targets that cannot be modeled as simple perfect electrical conductors (PEC) or to small "go-fast" boats in dynamic interaction with the sea surface.[75]

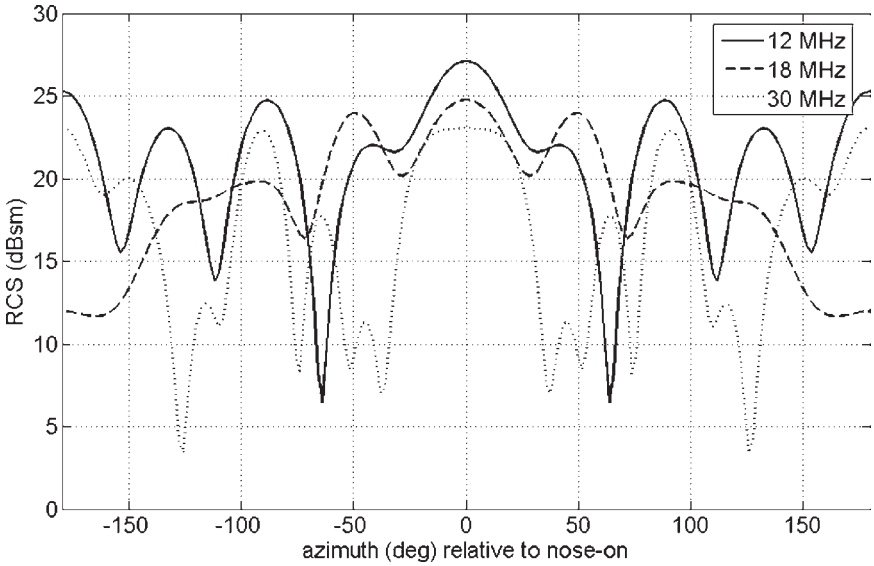


FIGURE 20.9 Monostatic RCS of the F-18 fighter aircraft at 12, 18, and 30 MHz, computed for HH polarization at a look-down angle of 5º

20.8 CLUTTER: ECHOES FROM THE ENVIRONMENT

Earth Surface Clutter. The geometry of skywave illumination ensures that target echoes will be immersed in returns from the Earth's surface, that is, clutter. In order to detect the targets, the properties of this clutter need to be understood so that the choice of frequency, waveform, and signal processing are compatible with the need to separate target echoes from clutter, and also so that the required dynamic range of the radar can be correctly specified.

Early HF radar experiments established that the strong ground clutter observed via skywave provided an indication of the physical characteristics of the illuminated terrestrial surface. Extensive observations made at the Naval Research Laboratory viewing alternately Atlantic Ocean areas and central United States areas indicated that, averaged over a wide area, sea clutter power levels were usually about an order of magnitude higher than those from an area of similar size in the central United States. Later observers noted extremely low backscatter from ice-covered areas of Greenland. These results are consistent with predicted scattering coefficient variations based on the topography and the electrical properties of the surfaces. Subsequently, observations over the Indian Ocean with the Jindalee radar, employing carefully calibrated transponders, revealed ~25 dB variations in the ocean scattering coefficient, depending on sea state.[76]

Land Clutter. Mapping of skywave backscatter from terrestrial (land) surfaces is of interest for two main reasons. First, a localized area of enhanced backscatter, such as a city in the central plains of the U.S. or a mountain rising from a tropical

rainforest, provides a geographical reference that can assist with the ever-present problem of coordinate registration[77] (uncertainties in the ray path traversed by the radar signals can lead to target positioning errors of over 100 km under some circumstances). Second, some regions experience strong seasonal variations in vegetation and soil moisture content, which may be reflected in measurable changes to scattering behavior. It is also important to understand the effect of the ground scattering coefficient and the topography when interpreting echoes from targets above the ground.

Sea Clutter. In contrast with the abrupt changes in scattering behavior that occur at coastlines or over complex terrain, the magnitude of the radar echo from the open ocean, that is, the scattering coefficient σ° tends to vary slowly with range and azimuth as a consequence of the scale lengths of typical oceanic meteorological systems and the response time of the ocean surface to varying wind stress. Moreover, much of the time, to a reasonable approximation, the echo power is proportional to the resolution cell area and can be used as an absolute amplitude reference when care is exercised. The reason for this is explained later in the discussion of radar oceanography.

Of far greater interest than the average magnitude of the sea echo is the wealth of information embedded in its doppler spectrum.[78] The waves on the sea surface introduce a complex modulation on the reflected radar signal, which is manifested in the signal's doppler spectrum. Estimation and interpretation of this modulation yields information on the time-varying sea surface geometry with major implications for ship detection as discussed below.

It transpires that a relatively simple model accounts for the observed properties of sea clutter with remarkable fidelity, provided the sea is not too rough. This model is based on two assumptions:

1. The sea surface can be represented to a good approximation as a superposition or spectrum $S(\vec{\kappa})$ of surface gravity waves satisfying the dispersion relation

$$\omega^2 = g\kappa \tanh(\kappa d) \tag{20.3}$$

where ω is the wave angular frequency, g is the acceleration due to gravity, κ is the wave number of the water wave, and d is the water depth. For deep water, this reduces to

$$\omega^2 = g\kappa \tag{20.4}$$

from which the water wave phase velocity can be written

$$v \equiv \frac{\omega}{\kappa} = \left[\frac{gL}{2\pi}\right]^{1/2} \tag{20.5}$$

where L is the wavelength of the water wave.

2. In terms of HF wavelengths, the sea surface can be regarded as only slightly rough, therefore admitting an approximate solution for the scattered field in the form of a perturbation series expansion in the parameter ka, where k is the radio wavenumber and a is a representative ocean wave amplitude. This approach, formulated by Rice[79] for the case of a static surface and extended by Barrick[80,81] to the case where the surface is evolving according to the dispersion relation (Eq. 20.4), leads to an equation for the doppler spectrum of the reflected radiowaves:

$$\sigma(\omega) = 2^6 \pi k_0^4 \sum_{m=\pm 1} S(-m(\vec{k}_{\text{scat}} - \vec{k}_{\text{inc}}))(\delta(\omega - \omega_B)$$

$$+ 2^6 \pi k_0^4 \sum_{m_1, m_2 = \pm 1} \iint \Gamma(m_1 \vec{\kappa}_1, m_2 \vec{\kappa}_2)^2 S(m_1 \vec{\kappa}_1) S(m_2 \vec{\kappa}_2) \tag{20.6}$$

$$\delta\left(\omega - m_1 \sqrt{g\kappa_1} - m_2 \sqrt{g\kappa_2}\right) d\vec{\kappa}_1 d\vec{\kappa}_2$$

where $\vec{k}_{\text{inc}}$ and $\vec{k}_{\text{scat}}$ are the incident and scattered radio wave vectors, $k_0 = |\vec{k}_{\text{inc}}|$, ω is the doppler frequency, $S(\vec{\kappa})$ is the sea directional wave spectrum, $\delta(.)$ is the Dirac delta function, and the Bragg frequency ω_B is given by $\omega_B = \sqrt{g|\vec{k}_{\text{scat}} - \vec{k}_{\text{inc}}|}$. The kernel $\Gamma(\vec{\kappa}_1, \vec{\kappa}_2)$ is discussed below. Eq. 20.6 reveals that the Barrick-Rice solution has a simple interpretation in terms of spatial resonance or Bragg scattering. Although the disorganized-looking ocean surface is represented as the Fourier sum of an infinite number of sinusoidal wave trains, each with its characteristic wavenumber and direction, the principal "first-order" contributions to the scattered field arise from only two ocean wave trains,[82] namely those whose wave vectors satisfy the relation

$$\vec{\kappa}_{\pm} = \pm (\vec{k}_{\text{scat}} - \vec{k}_{\text{inc}}) \tag{20.7}$$

For the simple case of backscatter at grazing incidence, the geometry for monostatic surface wave radar, $\vec{k}_{\text{scat}} = -\vec{k}_{\text{inc}}$, so these resonant ocean waves have a wavelength equal to one-half of the radar wavelength, with one solution corresponding to a wave directly approaching the radar and the other receding from it. The corresponding doppler shifts are those associated with the phase velocity of the resonant waves, that is

$$f_d = \pm \sqrt{\frac{g}{\pi \lambda}} = \pm \sqrt{\frac{gf}{\pi c}} \approx \pm 0.102 \sqrt{f(MHz)} \tag{20.8}$$

where the doppler shift f_d is in Hz; g is the gravitational acceleration (9.8 ms^{-2}), f is the radar frequency, and c the velocity of light.

With this insight, the second-order term in Eq. 20.6 can be interpreted as "double-bounce" processes involving Bragg scatter from first one and then another wave train, with the twice-scattered radiowave directed toward the receiver. Of course, there are, in principle, infinitely many pairs of wave trains that can satisfy this condition, hence, the integral. And there is another complication. The individual ocean wave trains are not completely independent—they interact weakly and produce evanescent nonlinear product waves that, while not freely propagating, change the geometry of the sea surface and contribute to the scattered field at second order, also via Bragg scattering. Thus, as shown first by Barrick,[81] the second-order scattering kernel is made up of electromagnetic and hydrodynamic terms, $\Gamma = \Gamma_{\text{EM}} + \Gamma_{\text{HYD}}$. The resulting piecewise continuous second-order doppler spectrum is usually some 20–30 dB weaker than the first-order Bragg peaks but, being spread in doppler, has the potential to mask ship echoes over much more of doppler space. Figure 20.10 is an example of the numerical evaluation of Eq. 20.6 for a specific ocean wave spectrum.

Polarization dependence arises through Γ. For a highly conducting medium such as seawater, viewed at typical HF radar geometries, the copolar surface scattering

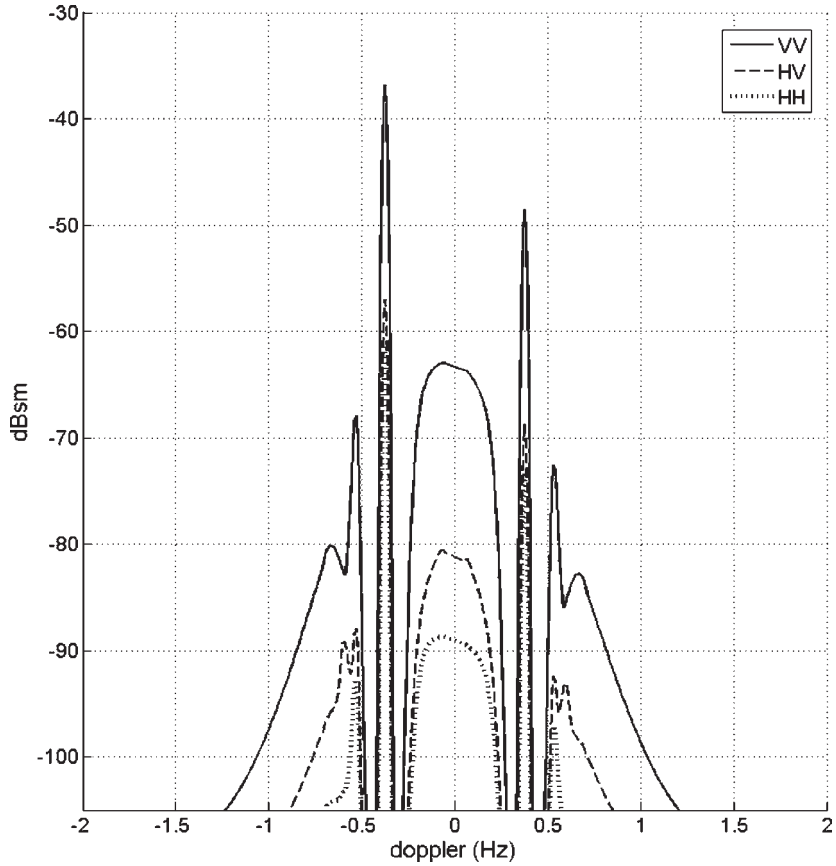


FIGURE 20.10 Computed doppler spectra of sea clutter for various polarizations; the radar frequency here is 15 MHz while the directional wave spectrum adopted is a Pierson-Moskowitz wave-number spectrum combined with a $\cos^4(\varphi/2)$ angular spreading pattern.

coefficient, or RCS per unit surface area, is much larger for vertical than for horizontal polarization, with the crosspolar scattering coefficients typically assuming intermediate values for moderate bistatic scattering geometries, $\sigma^\circ_{vv} \gg \sigma^\circ_{hv} \approx \sigma^\circ_{vh} > \sigma^\circ_{hh}$, as illustrated in Figure 20.10.

Eq. 20.6 is of fundamental importance to many HF radar applications, including ship detection, remote sensing (see following section), waveform selection, and other radar management functions. For example, Figure 20.11 shows a measured doppler spectrum and, superimposed, representative estimates of the magnitudes of the echoes that might be received from a number of different ship types for the same radar design and waveform parameters. Also shown, at left, are the speed bands within which the ship echoes would be obscured by the clutter. The ability to predict these obscured bands by means of Eq. 20.6 can be exploited for HF radar design and siting and for scheduling ship detection operations. Thus combining Eq. 20.6 with an ocean wave climatology for any region enables the statistical prediction of radar ship detection performance.[83]

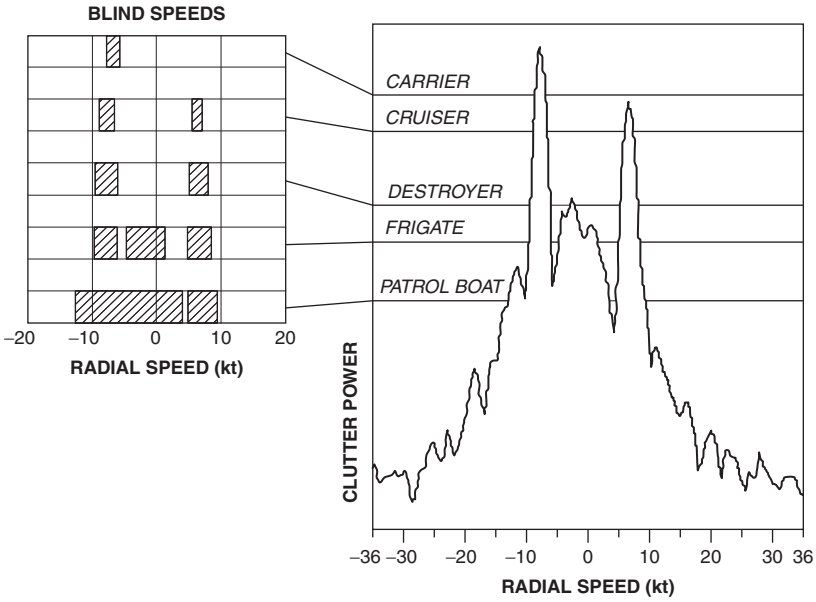


FIGURE 20.11 Schematic representation of blind speed bands in which ship echoes are obscured by sea clutter

Similarly, choice of the waveform frequency and bandwidth can be guided by calculation of the clutter power spectral density as a function of these parameters for any given sea state, taking into account any information about target course, speed, and RCS.

A very important operational consideration here is the potential for an adversary to exploit the sea clutter spectrum by contriving to place his ship echo at the Bragg frequency where it would be obscured by sea clutter. The adversary can achieve this by choosing any combination of course and speed that makes the velocity component toward the radar equal to the equivalent phase velocity of the Bragg peak. The radar operator must counter this by exploiting his understanding of the detailed structure of the spectrum and its variation with radar frequency to unmask the target.

Radar Oceanography. The existence of the relation expressed in Eq. 20.6 between the sea surface representation as a directional wave spectrum $S(\vec{\kappa})$ and the doppler spectrum measured by an HF radar provides an opportunity to determine the detailed state of the sea surface by remote sensing with skywave (or surface wave) radar. In order to extract sea parameters from radar doppler spectra, and to optimize the choice of radar parameters for the radar's surveillance missions, it is helpful to understand some elementary oceanography.

The ocean waves that contribute most to the HF radar returns have wavelengths in the range of 5–100 m; these waves are excited by the surface winds. If a wind blows at a constant velocity long enough and over sufficient fetch (the distance over which the wind is blowing), a steady-state condition will be achieved where the wind provides just enough energy to the waves to balance that lost by breaking and other dissipation mechanisms. Moreover, in this state of dynamic equilibrium, energy will be transferred

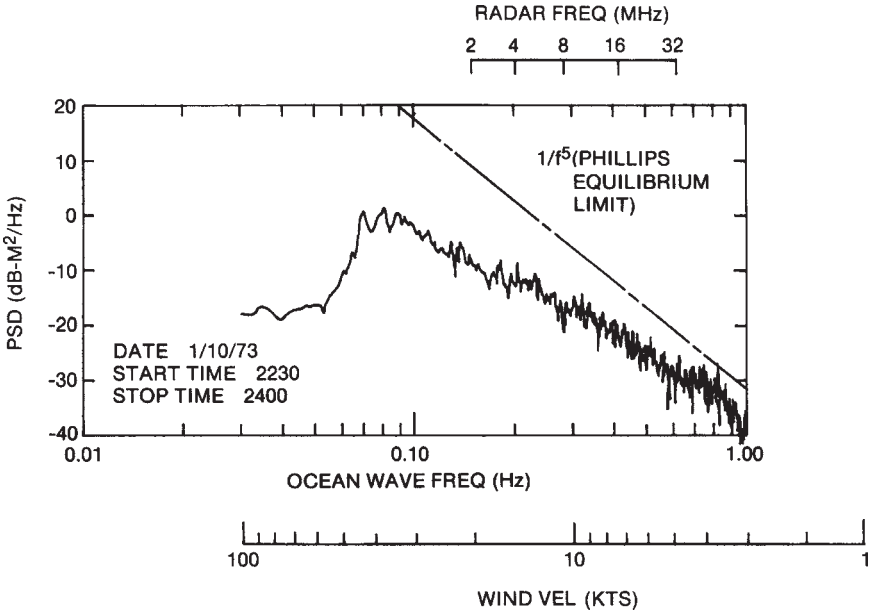


FIGURE 20.12 Nondirectional waveheight power spectral density, as measured with a wave buoy, showing power law behavior. The straight line is the original Phillips saturation asymptote, which is approached at high frequencies in this example. The scale marked "wind velocity" can be used to deduce that winds up to 40 kt have excited waves with frequencies as low as 0.08 Hz, but that either the duration or fetch, or both, have not been sufficient for full development. The scale across the top gives the radar frequency corresponding to resonant backscatter.

predominantly to waves whose velocities are fairly closely matched to the wind velocity and undergo redistribution across wavenumber space through the mechanism of nonlinear wave interactions to preserve the equilibrium spectral form.

There are a number of models that have been proposed to describe the equilibrium spectral form, and indeed, models that attempt the more ambitious task of modeling nonequilibrium spectra. Most of these models are based on experimental measurements, illustrated here by Figure 20.12, which shows an example of the frequency spectrum derived from wave buoy measurements. A common feature of such wave spectra is that waves of a given wavelength tend to reach a limiting spectral density, beyond which the processes of dissipation and nonlinear transfer of energy to other wavenumbers prevents further growth. This condition—known as saturation, or being fully developed—is reached at quite modest wind speeds for those waves responsible for first-order scatter at HF, that is 5–10 knots.

Of the various nondirectional ocean wave models reported in the literature, that of Pierson and Moskowitz has been most widely used by the radar community. They derived the following relation for a fully developed nondirectional spectrum based upon empirical data[84:]

$$F(\kappa) = \frac{\beta_e}{2\pi\kappa^4} \exp\left[-\nu\left(\frac{\kappa_c}{\kappa}\right)^2\right] \tag{20.9}$$

or equivalently, in terms of wave frequency,

$$\tilde{F}(\omega)=\frac{\beta_e g^2}{\omega^5}\exp\left[-\nu\left(\frac{g}{\omega u}\right)^4\right] \tag{20.10}$$

where

u = wind speed
κ_c = g/u^2
ν = 0.74 and $\beta_e = 0.0081$

The exponential term approximates the decay in the spectrum for wave speeds above the wind maximum velocity. Models that take account of the consequences of finite fetch and finite duration of the wind stress include the JONSWAP spectrum[85] and the model of Elfouhaily et al.[86]; such effects can lead to substantial changes in the HF doppler spectrum and hence target detectability, so these models should be used when applicable.

The decrease of wave spectral density with wavenumber is observed to lie close to a κ^{-4} power law, as adopted by most models, and by a natural coincidence, this is balanced by the k_0^4 factor in the first-order scattering coefficient in Eq. 20.6, so the resulting σ^0 is roughly independent of frequency over the range of frequencies for which the power law behavior is observed. An important consequence is that with such a reference, propagation path losses may be estimated.

To test this idea, the magnitude and variability of σ_{vv}° was examined with the San Clemente Island HF surface wave radar.[78] This radar facility had several valuable and unique features: a transmission path out over the open sea, multiple-frequency operation within a repetition period, calibrated antennas, known transmitter power, and ground truth in the form of ocean waveheight recordings. When looking into an approximately 20 knot wind, values of σ_{vv}° were found to be constant within a few decibels for operating frequencies where the ocean wave spectrum was approximately fully developed; these observations provided a confirmation of Barrick's first-order theory.[80] By using the antenna gain conventions stated earlier and assuming a semi-isotropic sea directional spectrum, the value of σ_{vv}° was calculated as –29 dB; the measured values were grouped between −7 and +3 dB of this value over a 5 to 20 MHz frequency span. This experiment provided the first direct measurements of the sea surface scattering coefficient.

Of course, when the Bragg resonant waves are not fully developed, the scattering coefficient will be proportionately less, as shown in Figure 20.13, compiled from data sets recorded looking upwind or downwind with the Jindalee radar.[76] Occasions when the wind speed was insufficient to arouse the Bragg resonant waves to saturation levels yielded scattering coefficient values up to 20 dB below the peak values.

Low scattering coefficient values can also arise because the angular spectrum of the wave system is not being sampled along a direction that presents the maximum amplitude component at the Bragg resonant wavenumber. The directional wave spectrum can be written as

$$S(\vec{k}) \equiv S(k,\phi) = F(k).G(\phi,k) \tag{20.11}$$

where $F(k)$ is the nondirectional spectrum,

$$F(k) = \int_0^{2\pi} S(k,\phi)d\phi$$

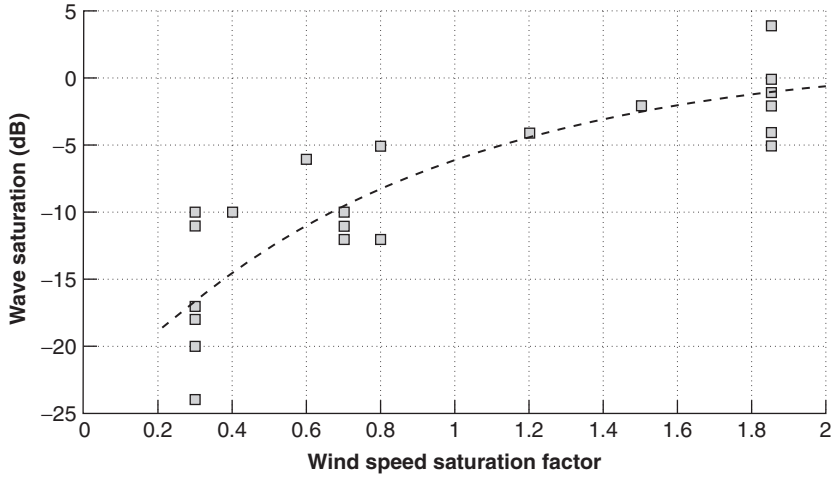


FIGURE 20.13 Calibrated measurements of the ratio of the measured scattering coefficient to the saturated sea scattering coefficient, plotted as a function of the ratio of wind speed to the phase speed of the Bragg-resonant waves for upwind/downwind observations

and $G(\phi,k)$ is the normalized angular spreading function that describes how the wave energy is distributed in azimuth,

$$\int_0^{2\pi} G(\phi,k)d\phi = 1$$

The non-zero wave spectrum is not confined to directions having a component parallel to the wind direction $\phi = 0$, i.e., $G(\phi,k)$ for $\frac{\pi}{2} \le \phi \le \frac{3\pi}{2}$, nor is it semi-isotropic, $G(\phi,k) = \frac{1}{\pi}$ for $\frac{-\pi}{2} \le \phi \le \frac{\pi}{2}$. In general, the angular spectrum is non-zero through 360°, with the spreading function depending on many variables, including the recent surface wind history. HF radar has sufficient sensitivity to measure the relative amplitude of the waves running against the wind, even though they may have a power spectral density (and hence an RCS) several orders of magnitude below those running with the wind. These upwind-propagating ocean waves are caused predominantly by third-order nonlinear wave-wave interactions, reflection processes, wave-current interactions, and propagation from neighboring regions with different wind stress. They are very important for remote sensing and impact strongly on target detection because the second-order scattering processes are heavily dependent on $G(\phi,k)$. They also serve as a sensitive indicator of backscatter coefficient estimation since, taking account of only the first-order scattered field,

$$\sigma^0 \propto [G(\phi - \phi_W, k) + G(\phi - \phi_W + \pi)]$$

where ϕ_W is the wind direction. To quantify this, if the wave angular spectrum model of Long and Trizna[87] is used, the maximum value of σ° for a saturated sea is –27 dB in the upwind or downwind direction (longitudinal sea) and only –39 dB in the crosswind direction (transverse sea). Figure 20.14 shows the scattering coefficient for this and some simple parametric angular spreading functions versus angle with respect to wind direction, as computed from the first-order contributions.

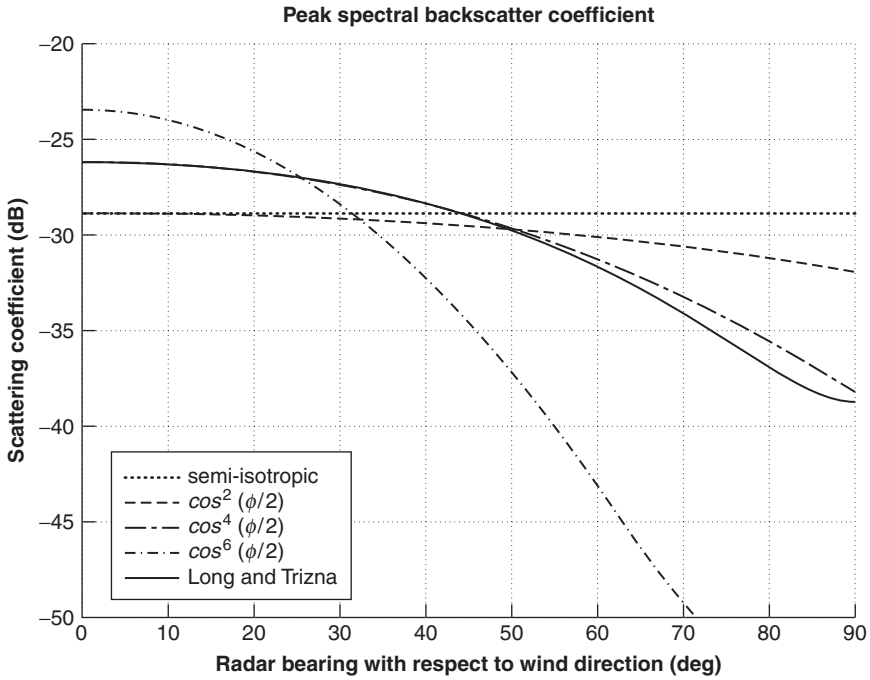


FIGURE 20.14 Variation of the peak backscatter coefficient as a function of wind direction relative to the radar look direction for various spreading functions, assuming a sea fully developed at the Bragg-resonant wave frequency

In summary, the sea echo power in a resolution cell (1) is generally the largest echo signal; (2) generally exists in the open ocean even in relative calm; (3) varies as the square of resonant waveheight, which is frequently saturated at the higher frequencies; and (4) varies with direction, being greatest for seas running toward or away from the radar. The doppler spectrum of the echo has sharp variations that require careful processing to preserve. The receiver and processor must be able to handle both the high-level signal due to this large RCS and those much smaller signals due to targets, especially when the latter are adjacent to the strongest clutter components. An HF radar must be designed to accommodate such clutter levels; even though, they will not exist all the time, or at any one time, over all areas, especially at the lower operating frequencies.

Estimation of Sea State, Wave Spectra, and Surface Winds. Techniques for extracting ocean wave field information from measured doppler spectra have been reported by numerous authors, working in almost all cases from the Barrick solution for the scattered field and dealing almost exclusively with HF surface wave radar data, uncorrupted by ionospheric propagation. Inversion of the relation in Eq. 20.6 to obtain an estimate of $S(\vec{\kappa})$ is mathematically nontrivial, requiring some additional assumptions to obtain a unique, stable solution. Some of these methods address the problem of full directional wave spectrum estimation,[88–90] whereas others propose estimators for integrated measures of sea roughness such as significant waveheight.[91]

Ocean currents can be determined when a zero doppler reference (such as an island) is in the radar footprint.[92] The NOAA Wavewatch website[93] provides access to archived maps of significant waveheight, dominant wave period, and other parameters; such information is extremely helpful when designing a radar for remote sensing applications.

In addition to providing information about the sea surface, HF radar can be used to infer surface wind speed and direction.[87,94] Wind direction is commonly estimated by taking the ratio of the first-order resonant Bragg peaks and employing an empirical relationship between this ratio and the wind direction relative to the radar look axis, as pioneered by Long and Trizna.[87] By scanning over a radar's coverage area, a map of inferred wind direction can be constructed; surface-wind direction maps are a routine byproduct of the Jindalee radar.[76]

While waveheight and wave spectrum estimates can, in principle, be extracted from higher-order features of skywave radar sea-echo spectra, a major difficulty arises due to the myriad forms of contamination and distortion introduced by the ionosphere. This has led to the development of numerous techniques for estimating and removing the various forms of signal corruption.[95–97] As an alternative, Trizna[98] and Pilon and Headrick[99] have reported a method for estimating σ° from simple measurements made directly on the corrupted radar echo spectrum. While this approach may be relatively insensitive to some forms of corruption, it is not applicable to seas far from equilibrium.

All the methods for estimating sea state or scattering coefficients require long coherent integration times, usually combined with noncoherent averaging of a number of CITs in order to achieve a distinct and stable spectrum. This type of radar operation will frequently be incompatible with other radar missions. But because the sea echo is generally a very large signal, it may be obtained with an adjunct oblique sounder operating in an appropriate radar mode.

Scattering from Meteor Trails and Other Irregularities in the Ionosphere. Clutter from ionospheric irregularities such as those mentioned in Section 20.4 can severely limit radar performance. Unlike the terrestrial or ocean environments where velocities of natural scatterers tend to be low, the ionosphere is home to phenomena with apparent speeds of 10^2–10^3 ms^{-1} and more, obscuring much of the relevant doppler domain where manmade targets might be found. Many of these scatterers have a transient existence, much less than the coherent integration time, so conventional signal processing causes their echoes to appear spread in doppler, as well as doppler shifted. The generic term *spread-doppler clutter* is used to cover all these phenomena where the scatterers responsible for the clutter do not have a well-defined doppler shift. The same term is used to describe the entirely different mechanism where clutter smearing in doppler occurs as a consequence of rapid variations in the propagation path rather than in the motion of the scatterer.

Meteors and their trails are the most ubiquitous source of transient echoes.[100–103] They display a reasonably well-defined diurnal and geographical distribution specific to any given radar site, but they can cause problems over a wide range extent because they can be illuminated by a variety of propagation modes,[19] as shown schematically in Figure 20.15. Normal incidence as shown yields the largest echo, but oblique incidence echoes and scattering from the meteor "head" are also observed. Figure 20.16 compares a simple predictive model with measurements, confirming that the observed behavior is understood and thus can be taken into account in radar design and operations.

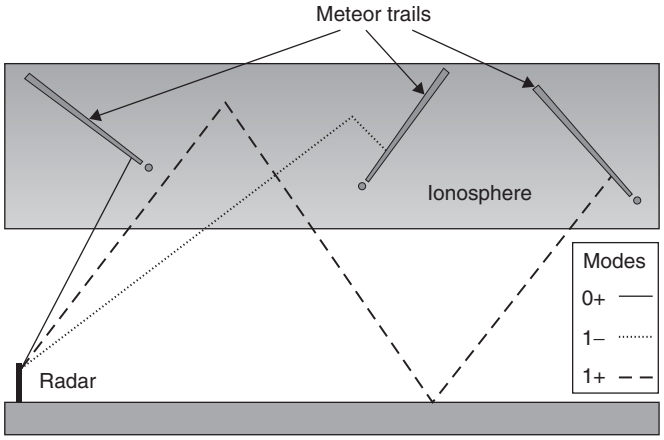


FIGURE 20.15 Meteor trail specular scattering geometry for several propagation modes

Meteors are usually classified as *sporadic*, occurring more or less randomly as the Earth moves around the sun, and *showers*, such as the Leonids and the Eta Aquarids, which occur on predictable dates where their orbits intersect the Earth's orbit.

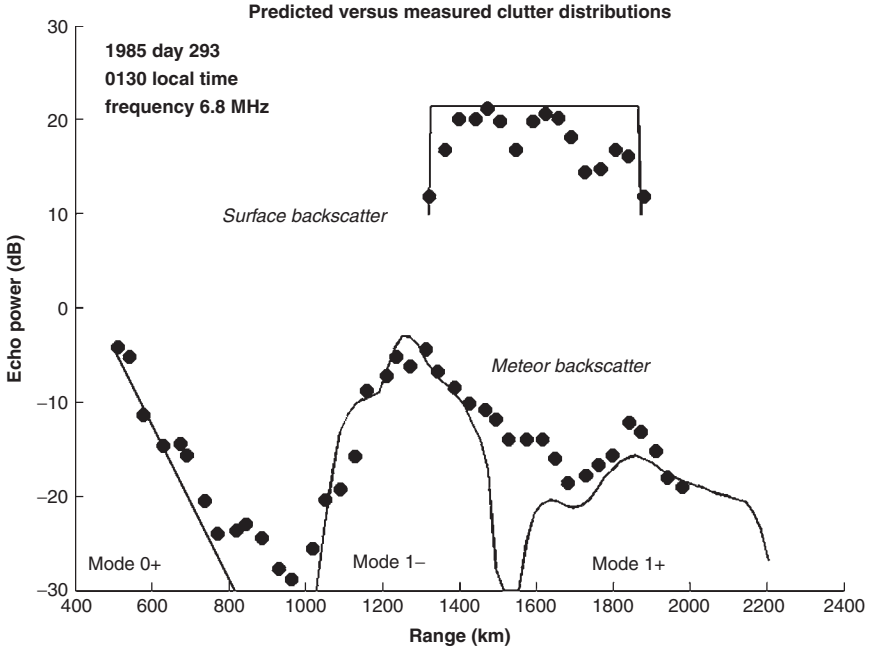


FIGURE 20.16 Predicted range dependence of meteor echo strength compared with measurements; ground clutter returns are also shown. The distribution of meteor radiants determines this variation.

On these occasions, the flux of meteors can be so high that they can cause serious obscuration of target echoes. As radars become more sensitive, they register smaller and smaller meteors, which are far more abundant, so at times, the flux of countless small meteor trails sets the effective detection threshold.

Meteor echo suppression can be attempted in the spatial domain if the receiving array has vertical directivity, but for most radars, the only option is rejection by signal processing, exploiting the transient nature of the echoes to detect and censor them in the time domain.

Echoes from the aurorae similarly involve transient scattering processes that appear in the radar data as highly doppler-spread echoes with the potential to obscure targets. Scattering from the auroral region has been studied extensively using the SuperDARN HF radar network initiated by Greenwald,[17] and an auroral echo-scattering model has been developed by Elkins[104]; this can be used to predict target obscuration when the transmission path is through the auroral region. Provided the radar is well removed from the auroral zone, the waveform can often be chosen so as to manipulate the clutter echoes within range-doppler space, unmasking targets hitherto obscured. Often several waveforms with different repetition frequencies are interleaved to achieve this. For radars close to the auroral zone, the options are generally more limited.

Ionospheric irregularities that scatter back to the radar receiver occur much more often at night than by day at any latitude. Sensible siting of the radar, ensuring good front-to-back ratios on the antennas, employing vertical nulling if available, and using adaptive signal processing techniques (or at least maintaining low receive array sidelobes) are all effective tools for mitigating auroral, meteor, and other ionospheric clutter.

20.9 NOISE, INTERFERENCE, AND SPECTRUM OCCUPANCY

In the HF band, the average noise power spectral density at mid-band, near 15 MHz, say, may exceed –150 dBW/Hz and will generally exceed –175 dBW/Hz, compared with typical receiver internal noise spectral densities of perhaps –195 dBW/Hz. Thus, unlike the microwave radar case, external noise is almost always dominant. This has fundamental implications for receiving system design and signal processing. Another critical issue is the observed systematic variation of the external noise level, which has a direct impact on radar performance.

The major source of quasi-continuum background noise at the lower frequencies is lightning discharges ionospherically propagated from all over the world *(sferics).* At the high-end of the band, extraterrestrial or galactic noise may be greater than that due to sferics. Receive sites in an area of extensive electrical equipment use can find anthropogenic noise dominant. But most importantly, the HF band is densely occupied by other users, especially powerful HF broadcasters relying on the prevailing frequency window for satisfactory propagation. Even out-of-band signal levels are a consideration in receiver front-end design, where it is common to have bandwidths much wider than that of the radar signal. There are a large number of broadcast stations that have 500-kW transmitters and antennas with more than 20-dB gain. Measurements made on the middle Atlantic coast of the United States show HF broadcast-band signals with strengths of 5 to 10 mV/m. These ambient levels must be accommodated in receiver design because a wideband front end is desirable for rapid and frequent frequency changes.

The practice in allocations for HF radar operation is to permit use of broad bands of the spectrum with a requirement to cause no discernible interference to an existing service and to provide a lockout feature for channels that need protection. Thus, an integral part of an HF radar is a channel occupancy analyzer that provides a real-time description of spectrum availability, as illustrated in Figures 20.17 and 20.18.

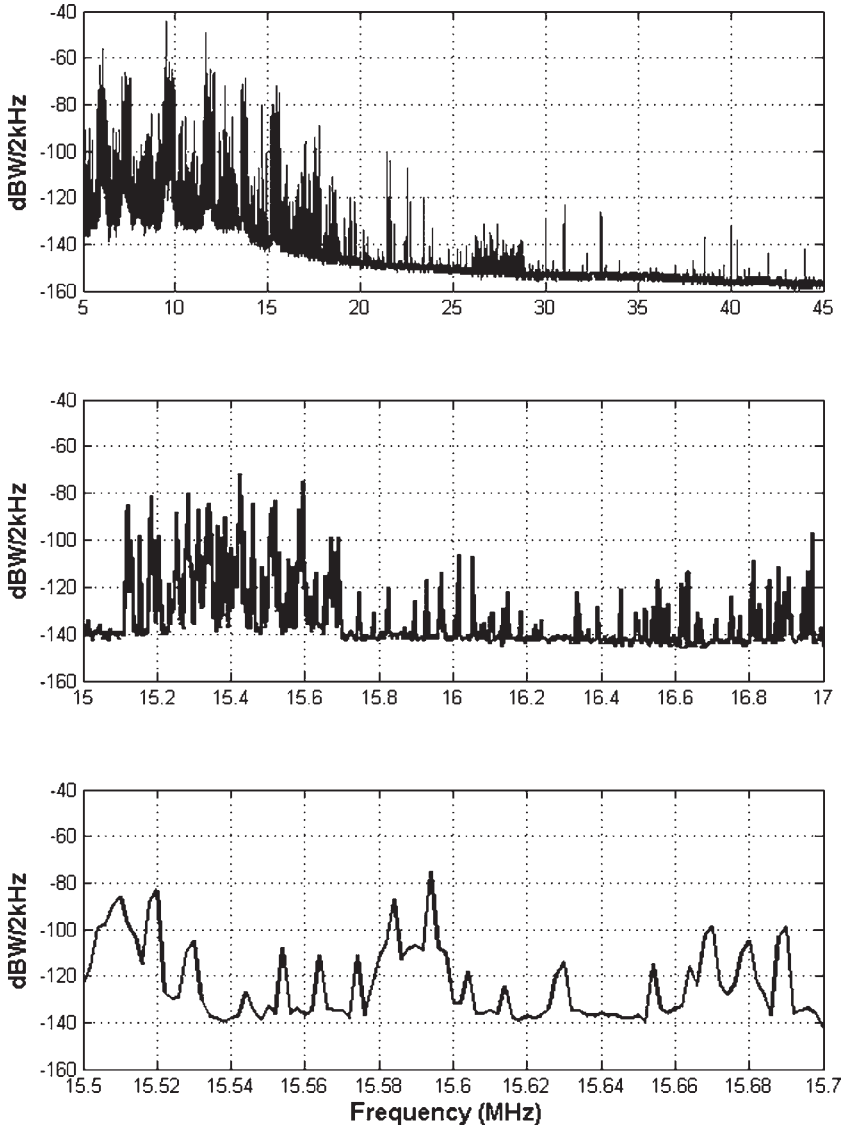


FIGURE 20.17 A snapshot of the HF spectrum at noon in summer and at a low sunspot number, zooming in progressively to show a section of the spectrum where a 20 kHz clear channel is evident, centered on 15.640 MHz. The data was recorded at latitude 23.6 S, longitude 133.1 E.

Almost without exception, HF radars have operated on this principle of noninterference, making do with "clear" channels between other users.

Patterns in Spectrum Occupancy. When the HF band is scanned with a spectrum analyzer, it can be seen that the gross features of occupancy at any particular hour are remarkably stationary over the days of a season. This is due to broadcast stations, fixed-service point-to-point transmitters, and many other spectrum users having regular schedules, as is evident from Figure 20.18, which plots the power spectral density at 2kHz resolution for high and low solar activity, summer and winter, day and night.

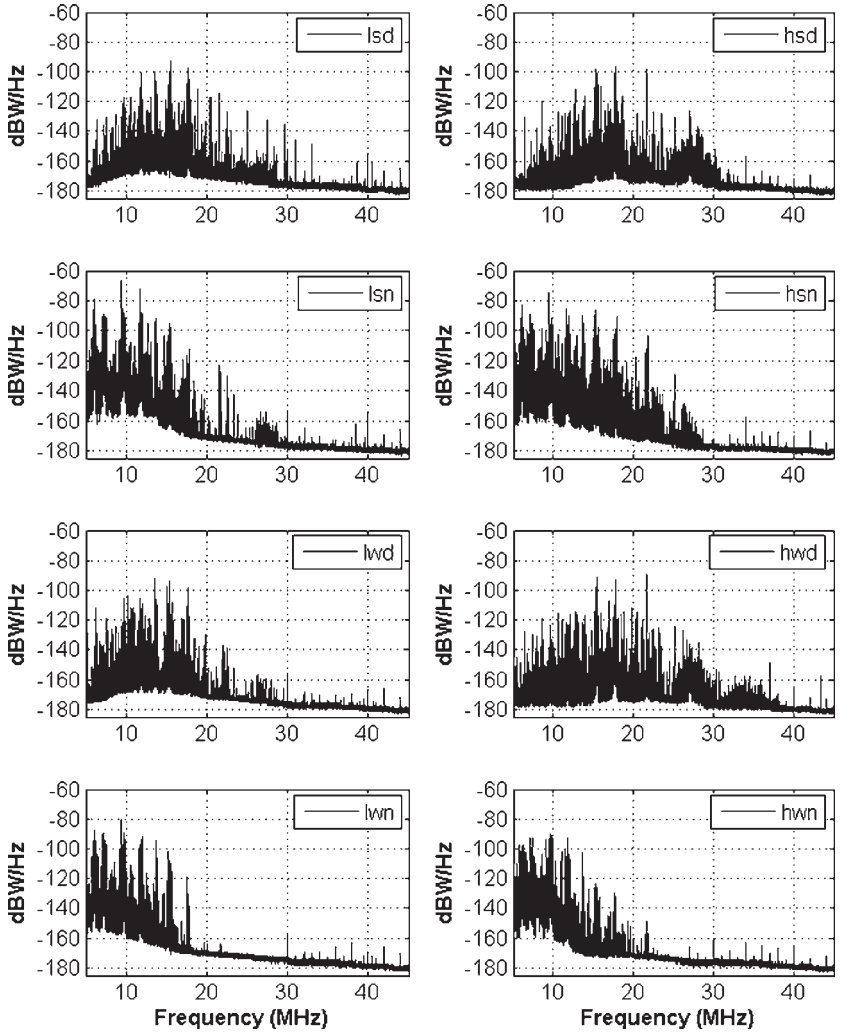


FIGURE 20.18 HF activity over the 5–45 MHz for low and high sunspot numbers (l, h), summer and winter (s, w), and day and night (d, n), as measured at latitude 23.6 S, longitude 133.1 E. The data is taken from Julian days 180 and 360 of 2000 and 2005.

As remarked earlier, the maximum frequency that will reflect energy back to the Earth during the day may be more than twice that at night; therefore, the occupancy tends to be denser at night than during the day—a problem compounded by the lower absorption and, hence, reception of more distant signals.

Looked at over longer timescales, various patterns and trends emerge. Most obviously, the 11-year solar cycle forces changes to spectrum usage and the density of users, with a resultant impact on HF radar channel selection. Another trend that has become apparent in recent years is the gradual reduction in HF users as services move to satellite communications, microwave links, fiber optics, and other media. Nevertheless, the increased number of HF radars has led to a new challenge: inter-radar interference and the need for frequency arbitration.

Noise Models. The widely used reference on noise is the International Radio Consultative Committee (CCIR) Report 322.[105] This report is based upon measurements made at 16 locations throughout the world. The measurement and data analysis were performed to exclude individual collection site local thunderstorm contributions. Spaulding and Washburn[106] added data from the former U.S.S.R. for two revised CCIR reports.[105] Noise-level medians as a function of frequency are given in the form of worldwide maps by season and 4-h time blocks. Lucas and Harper[107] have provided a numerical representation of CCIR Report 322-1, which is useful for computer computations, and this has been revised by adding the work of Spaulding and Washburn. The maps of median values are accompanied by decile values to indicate distributions over days of the season. (CCIR Report 322-3 (1988) has a significant discrepancy, as pointed out by Sailors,[108] so it should be used with caution.) These noise maps provide the level that an omnidirectional antenna would receive. Even though assuming isotropic CCIR noise has limitations, it does provide a reference level for initial radar design. A number of operational and experimental HF systems have accumulated their own noise databases and compared them with CCIR model data; the report of Northey and Whitham gives a detailed analysis.[109]

An HF radar is generally designed to take advantage of what the environment permits; that is, the receiver noise figure should be good enough to make environmental noise the limitation.

An example of the CCIR Report 322 data will now be discussed. Figure 20.19 was drawn from Lucas and Harper.[107] Noise power in a 1-Hz band relative to 1 W (dBW) is given as a function of frequency for three different sources of noise: galactic, atmospheric, and anthropogenic. The practice in use is to select the largest. This is a winter daytime example at a United States east coast location. The three straight lines are estimates of anthropogenic noise for three different types of sites. The shape of the anthropogenic curves is described by the equations

$$\begin{aligned} N_o &= -136 - 12.6 \ln (f/3) \quad \text{residential} \\ N_o &= -148 - 12.6 \ln (f/3) \quad \text{rural} \\ N_o &= -164 - 12.6 \ln (f/3) \quad \text{remote} \end{aligned} \tag{20.12}$$

where the frequency f is in megahertz and ln indicates the natural logarithm.

These frequency trends approximate many measurements of human-made noise but ideally the curve would be based on measurements at the particular radar site. The galactic-noise curve should be selected when it is the largest and when there is a path through the ionosphere; the path will not exist for the lower operating frequencies in the daytime. The atmospheric noise rises from low frequencies to about 12 MHz and

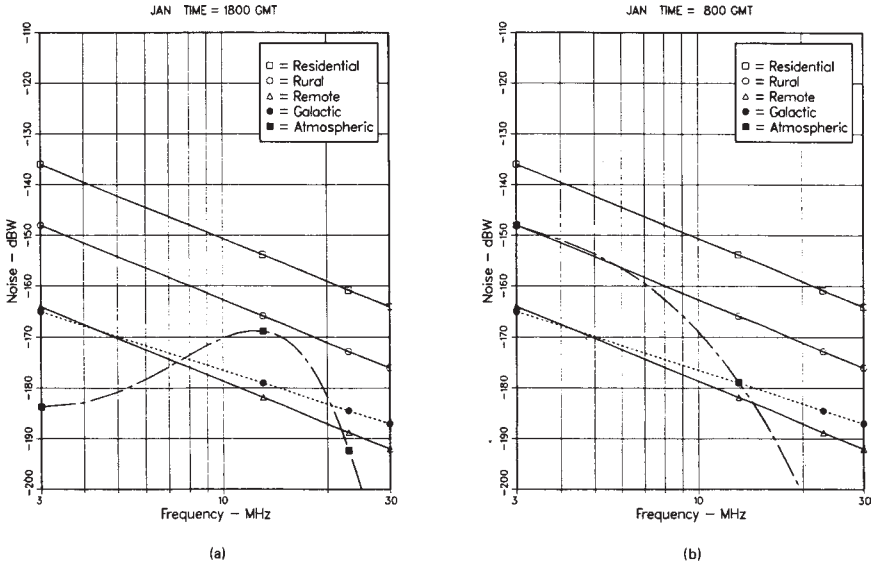


FIGURE 20.19 Noise power per hertz is given for 38.65' north latitude and 76.53' west longitude in winter: (*a*) 1800 UTC is given as a daytime example and (*b*) 0800 UTC is given as a nighttime example.

then rapidly falls. Figure 20.19*b* is for nighttime. All the curves are the same as in Figure 20.19*a* except for atmospheric noise. At 10 MHz, the night and day levels are the same; below 10 MHz, the noise decreases with decreasing frequency in daytime and increases at night. Above 10 MHz, daytime levels are greater than those at night. These effects can be partially explained by the very lossy long-range paths in day that attenuate the long-range noise at the lower frequencies and by there being few or no skywave paths to noise terrestrial sources at the higher frequencies at night. In general, nighttime noise will be greater than daytime noise for skywave illumination of a selected range. This is evident in Figure 20.20, recorded at time zone UT + 9½. The general trends of atmospheric noise in other seasons are similar to those in winter. However, there can be large differences in levels at other locations on the Earth.

For more detailed analysis and system optimization, it is no longer acceptable to treat the noise as an isotropic field. Strong azimuth- and elevation-angle dependence of the sferics field is inevitable: examination of maps produced by satellites [110] indicates that tropical rain forests and other regions of concentrated thunderstorm activity are major sources of noise, and these regions are connected with a given radar site by the constraints of skywave propagation. Coleman[111] combined the thunderstorm activity maps of Kotaki[110] with numerical ray-tracing and modeled antenna patterns to demonstrate that the directional variability of noise, coupled with the directional characteristics of different antennas, can lead to marked changes in noise outcomes and hence in optimal radar design.

Another important issue is the behavior of the external noise field as a function of time, that is, within the coherent integration interval since this impacts strongly on signal processing for interference rejection.[112] Other effects that can influence radar performance are sometimes mistaken for the *additive* noise discussed above. One of

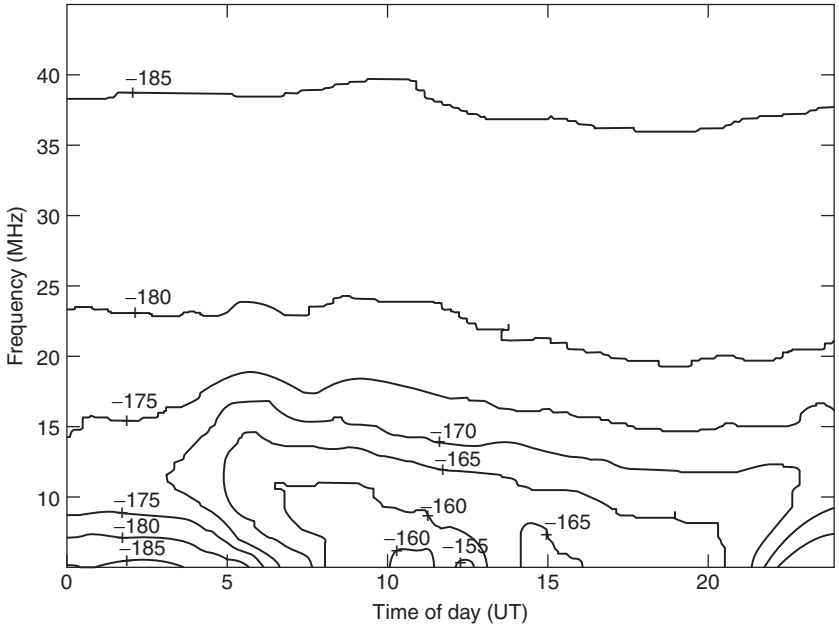


FIGURE 20.20 Diurnal variation of HF noise, measured in dBW/Hz, at latitude 23.6 S, longitude 133.1 E, 5 November 1995

these is the spread doppler clutter discussed in Section 20.8, sometimes referred to as *active* or *multiplicative noise*. The occurrence of this type of clutter is greater at night and is much more prevalent in the auroral zones and around the magnetic equator.

20.10 THE RECEIVING SYSTEM

The receiving system is defined here to embrace only the receiving antenna array and the receivers that convert the antenna outputs to discrete time series, usually at baseband. For ease of reference, the conventional signal processing stages responsible for transforming the receiver outputs into standard radar products are discussed in Section 20.11, together with more specialized techniques.

Antennas. From the reception viewpoint, it is desirable to have fine azimuthal resolution for several reasons, including (i) to improve target location accuracy and tracking performance, (ii) for detailed clutter mapping, and (iii) to reduce the clutter amplitude levels to values permitted by system dynamic range and slow-target detection requirements. As mentioned earlier, most HF skywave radars employ a broad transmit beam to illuminate a zone of interest, and then process the echoes via a number of much higher resolution simultaneous "finger" beams. It is well known that the classical resolution of an array improves linearly with aperture up to some limiting value determined by the environment. Horizontal apertures of 2–3 km and even greater

have been shown to support spatially coherent processing.[113,114] Hence, the receiving array apertures of skywave radars may range from ~ 0.3 km for systems concerned only with aircraft and ballistic missiles to more than 4 km in systems designed for ship detection and tracking. Taking 3 km as a receive aperture, conventional beamwidth at 15 MHz is ~ 0.5°, so 20 simultaneous beams span ~ 10°, which sets the required transmit aperture in this example at 120–150 m, depending on the variation in gain that can be tolerated across the set of receive beams. The most versatile beamforming techniques rely on having a receiver for each array element, but as the number of elements may approach 500, the cost of such a solution may be prohibitive. It is then advantageous to configure the array elements into subarrays, possibly overlapped or sharing elements, with one receiver per subarray. For instance, the original Jindalee radar grouped 462 elements spanning 2.766 km into 32 overlapped subarrays.[115] While this constrains the resultant beams to lie within the subarray angular response pattern, it has the advantage of reducing receiver dynamic range requirements by suppressing interference from other angular sectors.

Linear arrays provide the most economical route to high spatial resolution in azimuth, but the existence of multiple paths for signal reception, typically via the E, F1, and F2 layers, has motivated some designers to employ two-dimensional arrays, horizontally or, at greater cost, vertically deployed. The relative merits of these designs can be measured only with respect to the priorities accorded to the various missions assigned to a given radar.

The choice of receiving antenna element has traditionally been based on the precept that, at HF, the external noise almost always exceeds the internal noise by a substantial margin. On this logic, improving antenna efficiency increases the output external noise and interference amplitude at the same rate that it improves the wanted signals, thereby gaining no advantage in SNR. Selection of antenna element type, such as monopoles, dipoles, Beverage antennas, phased endfire rows of monopoles, or biconical antennas, for instance, can then be based on frequency response over the anticipated band of interest and suitability for the chosen array geometry, as well as terrain constraints such as the soil conductivity. Studies have shown that this argument is not necessarily valid when advanced adaptive spatial and temporal processing techniques are employed because interference rejection efficacy is enhanced by higher interference-to-internal noise ratios.

It has been widely argued that the inevitability of time-varying polarization transformation in the course of ionospheric propagation greatly reduces the possible utility of being able to measure the polarization state of the signals arriving at the receiving array. This is an open question at this time, though experiments aimed at assessing skywave radar polarimetry are underway.[116]

Receivers. There are many demands on the receivers for OTH radar, including high dynamic range, linearity, wide bandwidth, and uniformity between receivers when used in multireceiver systems. For most civil aircraft and ships, target radar cross section (RCS) at HF is roughly of the same order as the microwave RCS, that is, ~10–20 dBsm for aircraft and ~30–50 dBsm for ships, but the range is 10–100 times greater, so the extra loss associated with R^{-4} is in the range 40–80 dB. Moreover, each target echo is immersed in clutter from the illuminated footprint, which may have an area of many thousands of square kilometers. Further, the HF signal environment includes (one-way) transmissions from powerful radio stations around the world, as discussed in the previous section. Imperfections in the receiver result in some of this

noise and clutter energy being superimposed on the wanted radar echoes, either additively or multiplicatively. Hence, careful attention to receiver design is imperative if the radar designer wishes to avoid self-inflicted performance limitations.[117]

Attempts to reduce contamination from external broadcast signals by inserting narrow-band filters at the receiver front-end sacrifice the high agility that is needed when the radar is changing frequency, typically by several MHz, second by second, as it jumps between tasks. There are also penalties from (i) filter switching time, (ii) settling time, (iii) distortion caused by group delay dispersion, and (iv) reduced reliability when there are hundreds of receivers. Further, each channel will need to account for the gain and phase variation for each filter, increasing the overheads on band switching. It is better to zero in on the bandwidth of interest by nonswitched filters later in the receiver, using a variable frequency local oscillator to position the desired subband(s) over the selective filters. Of course the switched LO can also suffer from imperfections, but only one local oscillator is needed, as opposed to hundreds of receivers. Whichever design path is followed, the demands on receiver linearity and spurious-free dynamic range are extreme.

There are five dominant mechanisms known to degrade HF radar receivers: the nonlinear processes of analog-to-digital conversion, out-of-band inter-modulation (IMD), cross-modulation, and in-band inter-modulation, and the pseudo-linear process of reciprocal mixing.

Analog-to-digital Conversion. Analog-to-digital conversion involves two stages—sampling and quantization—each with potential for distorting the received signal. The received signals must be sampled with sufficient precision and uniformity to preserve the inherent spectral content across the dynamic range spanned by the signal components—target echoes, clutter, and external noise—after taking account of the artifacts introduced by quantization and timing jitter, especially in multi-receiver systems.[118,119]

Out-of-band Inter-modulation. Out-of-band IMD arises from nonlinear mixing of two (or more) strong interferers such as broadcast stations, where the powerful signals enter the front end of the receiving system and generate IMD products within the radar signal bandwidth before they, the original interferers, are rejected by selective filtering.[120,121]

Cross-modulation. Cross-modulation involves nonlinear mixing of a strong interferer with the received radar echoes, transferring the interferer modulation onto the radar signal.[120,121]

In-band Inter-modulation. A single resolution cell in the radar footprint may have an area of 50–500 square kilometers, and there may be hundreds of cells in the radar footprint, so using representative values of target RCS (Figure 20.9) and surface scattering coefficient σ° (Figure 20.14), the signal-to-clutter ratio in an OTH radar receiver may be as low as –80 dB for systems employing FM-CW waveforms. Nonlinearity at any stage has the potential to mask target echoes by mixing the clutter echo with itself, generating IMD products that extend beyond the intrinsic doppler band of the clutter. Unlike the out-of-band IMD, in-band IMD can occur at any stage through the receiver, and hence even second-order products may cause problems.[122]

Reciprocal Mixing. Analog receivers of heterodyne design generally involve a number of local oscillators for signal mixing and perhaps a waveform generator used to implement a matched filter. These ancillary sources are inevitably of finite spectral purity, with a phase noise floor that may extend over a wide band of frequencies, albeit at a very low level. Any powerful interfering signals entering the first mixer stage of the receiver will combine with the phase noise floor and potentially generate products in the radar signal bandwidth.[123,124] In the case of digital receivers, sampling noise will have an equivalent effect.

Digital Receiver Technology. HF receivers can employ direct digital conversion at RF, avoiding some of the limitations of analog devices. Issues such as reciprocal mixing are still present, though in slightly modified form. Comparisons with analog receivers have demonstrated that little or no penalty is incurred with such designs if a band preselector filter is installed at the front-end. On the other hand, there are major advantages to be had, such as simultaneous reception of several radar signals on different frequencies, so a single receiver can service more than one radar transmitter.

Calibration. The conventional beamforming process in an ideal multi-channel receiving system should deliver a single output, assigned to the correct direction, for a plane wave incident on the antenna array. Inevitable variations in gain and phase arising from small antenna positioning errors, ground screen inhomogeneities, differences between preamplifiers, mutual coupling, cable mismatch, thermal and other variations of cable characteristics, and all the analog stages of the receiver result in distortion of the beam shape (and hence degraded radar resolution), elevated sidelobes (and hence vulnerability to clutter and interference), pointing errors (and hence increased tracking errors), and interferer wavefront geometry perturbations (and hence wasted degrees of freedom in adaptive beamforming).

To mitigate these effects, HF radars must employ sophisticated calibration schemes. Several approaches have been tried:

- Use of an external radiating element in the near-field in front of the array. This scheme is vulnerable to variations in the soil electrical and magnetic properties, such as those resulting from seasonal changes in soil moisture levels.
- Injection of a calibration waveform at the antennas, or at the receiver inputs behind the antenna array, by means of an independent "open loop" signal feeder network.[125] This scheme cannot calibrate the antennas and initial feeders; it applies only from the point where the signal is injected.
- Use of a distant radiating source that illuminates the array via skywave or even a discrete target echo. This scheme presumes that the arriving signal wavefronts are essentially planar or smooth after ionospheric reflection, which is not always the case.[126]
- Joint analysis of multiple discrete meteor echoes.[127,128] This scheme is appealing but relies on having enough identifiable discrete echoes. For a 32-element array it may be generally viable, but less frequently for a 500-element array.
- Receiver and plane wave rejection tests applied to injected broadband noise.[129] This approach provides useful metrics of calibration performance and relative performance, but does so only downstream from the receiver inputs, as with (ii) above.

Each of these approaches has its advantages and drawbacks in the context of conventional (Fourier) beamforming. When adaptive processing techniques are being employed, it is also important that the effective look direction matches the array steering vector or else cancellation of the wanted signal will occur.

20.11 SIGNAL PROCESSING AND TRACKING

Signal Analysis and Target Detection. The objective of signal processing is to detect and characterize echoes from scatterers of interest, either discrete (aircraft or ships) or extended (the sea surface), and this is customarily achieved by decomposing the time series data from the receivers into the natural radar domain dimensions of group range (based on time delay), direction of arrival (beam space), and doppler frequency, hopefully separating the echoes of interest from unwanted clutter and noise. The standard tool for this decomposition is the FFT, at least in operational skywave radars, in part because it is computationally quite feasible to analyze the incoming signal into typically ~10^2 range bins, $10^1 - 10^2$ beams, and ~10^2 doppler cells, in perhaps ~10^0 seconds, using general purpose computing hardware. Thus, the FFT (or DFT for short transforms) is commonly used for the three dimensions of analysis. Alternative analysis techniques have been implemented in some systems for applications such as detection of accelerating targets,[130–132] the detection of harmonically related signals,[133] and when high doppler resolution is required but only short coherent integration times can be accommodated in the radar timeline.[134]

Some important processor design considerations emerge from an analysis of high quality HF radar data, as illustrated by the following example. Figure 20.21 shows a sequence of sea clutter doppler spectra from the AN/FPS-118 radar, presented as received power amplitude versus doppler frequency. To generate this plot, 15 power spectra from a single range bin were computed from nonoverlapping time intervals, and then grouped into blocks of five and averaged noncoherently. The waveform repetition frequency (WRF) in this case was 20 Hz, with a coherent integration time of 12.8 s, resulting in a nominal doppler filter bandwidth of 0.08 Hz. Noise (N) samples were taken from the maximum doppler bin, target samples (T) on the target peak, and Bragg line amplitudes (A) and (R) from the clutter peaks corresponding to the approaching and receding resonant ocean waves, as described in Section 20.8. N, T, A, and R are plotted in Figure 20.22*a*. The sub-clutter visibility (SCV), widely used in HF radar as a measure of sensitivity and defined as R/N, is 76 dB in this example. (In microwave radar, the term *sub-clutter* visibility is sensibly defined as the ratio by which the target echo power may be weaker than the clutter power and still be detected. In the HF radar literature, historically the detection threshold component has not been included, so SCV is essentially the clutter-to-noise ratio.)

For a representative effective clutter RCS of 65 dBsm per resolution cell, the target RCS can be estimated as

$$\text{RCS} = 65 - R(\text{dB}) + T(\text{dB}) = 65 - 76 + 34 = 23 \text{ dBsm}$$

and supposing that the SNR required for registering a detection is 15 dB, the minimum detectable RCS (MDRCS) in this example can be computed as

$$\text{MDRCS} = 65 - R + N + 15 = 65 - 76 + 15 = 4 \text{ dBsm}$$

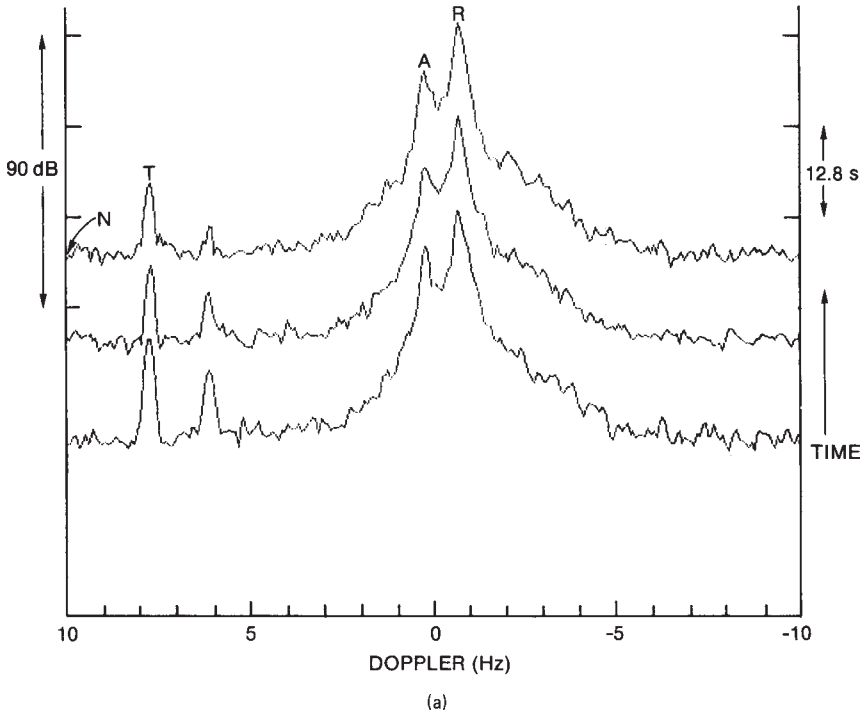


FIGURE 20.21 Noncoherently averaged power spectra from three contiguous time intervals, plotted with an offset for clarity. The target lies well outside the clutter region where the Bragg lines are clearly visible, so the signal-to-noise ratio, not the signal-to-clutter ratio, determines detectability. Note the ~10 dB variation in target echo strength due to changes in the ionosphere over 128 seconds.

This is a quite low RCS at HF, indicating that when conditions are favorable, very small targets are potentially detectable.

The experimental demonstration that, on occasion, the environment supports ~80 dB sub-clutter visibility obliges the radar designer to ensure that the receiving system and signal processing operations do not unwittingly degrade the radar performance. Some appropriate design considerations are

- An analog-to-digital (A/D) converter of at least 16-bit precision is in order for HF radars with high power-gain products, $P_{\text{av}}G_TG_R$.
- Receiving apertures, waveform bandwidths, and coherent integration times should provide enough samples and high enough sampling rates to resolve unambiguously distinct features of the clutter spectrum where such resolution impacts on target detectability or the extraction of important information.
- The window functions used to control leakage during conventional spectrum analysis must have sufficiently low sidelobes when transforms are to be applied directly to high dynamic range data.

The basic steps of range, azimuth, and doppler analysis are not the only stages of signal processing. As explained in Section 20.4, HF skywave radar signals are subject

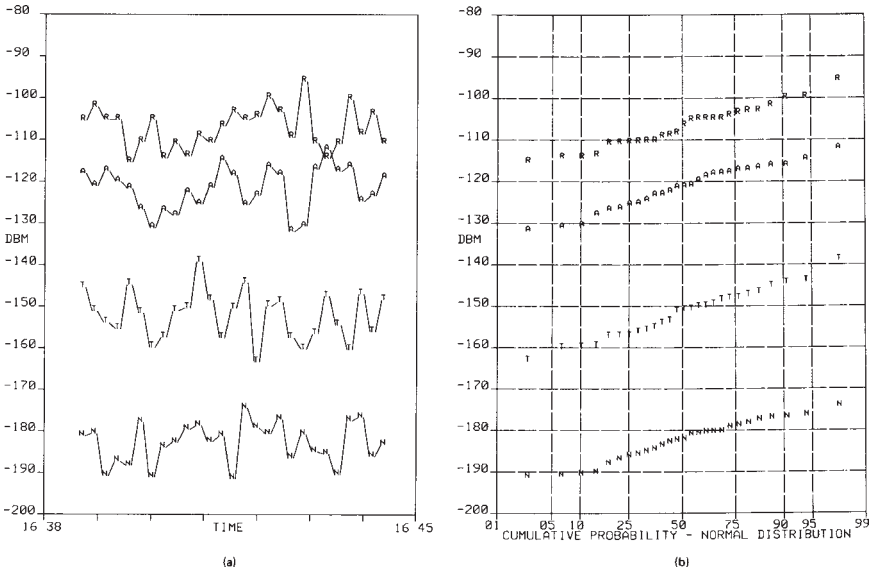


FIGURE 20.22 (*a*) Temporal fading and (*b*) distribution properties of the target, clutter peaks, and noise features in the doppler spectrum of Figure 20.21, computed from an extended data sequence

to various forms of contamination and distortion, many ionospherically induced, so methods to deal with these have long been part of the radar signal processing toolbox.[135,136] Moreover, the need to deal with these deleterious effects has become more pressing as the achievable dynamic range of receiving systems has increased, revealing a greater variety of signal distortion mechanisms. Hence, in addition to performing the basic decomposition described above, and calibrating the receiver chain for the current waveform, the signal processing stage may be tasked with a number of signal conditioning operations. Signal conditioning here refers to filtering and scaling processes aimed at removing contamination and distortion that, if left on the signal, would degrade the results of the primary processing operations, such as doppler analysis, and be more difficult to remove after that primary processing. Typically, there are (optional) processing algorithms available to remove impulsive noise originating from lightning, transient echoes from meteors, and field-aligned irregularities in the ionosphere, strong directional interference, and clutter echoes from beyond the maximum unambiguous range. In addition, the radar designer now has the option to compensate for signal path fluctuations in both spatial and temporal domains, within limits, by employing processing schemes that diagnose the nature of the signal corruption on a first pass and then reprocess with algorithms that compensate for the observed corrupting mechanisms.[96,97] The signal processing stage may also be tasked with extracting environmental information from the radar echoes, as mentioned in Section 20.8. This includes remote sensing of ocean winds and sea state, land-sea mapping for coastline recognition to assist with coordinate registration, measurements of the ionosphere for assimilation into ionospheric models, extraction of echoes from remote beacons deployed to assist with coordinate registration and calibration, and many other byproducts of the basic detection mission.

At present, the most powerful tools for dealing with additive interference, as well as some classes of spread-doppler clutter, are the techniques of adaptive processing. Originally, these techniques were developed for airborne microwave radars, but they have been shown to have wide applicability in HF radar.[135–138] The primary domain where major performance gains are realized is spatial analysis, that is, beamforming. In this context, a "snapshot" of the outputs from the all the receivers across the array is taken and used to compute a set of complex-valued weights for each desired beam direction; these are then applied to the receiver outputs before summing to form the adapted beams. The reason for the efficacy of adaptivity here is that, as remarked previously, HF radars are almost always external noise–limited. The azimuthal distribution of noise from thunderstorms, industrial sites and other sources is far from uniformly distributed. Furthermore, even so-called clear channels selected for radar operations are contaminated with directional noise of natural or industrial origin, albeit at a much lower level. Conventional FFT-type beamforming makes no allowance for this, so a lot of noise energy leaks into each computed beam through those of its (regularly spaced) sidelobes that are pointing toward strong noise sources. Spatially adaptive processing (SAP) reduces this problem by adjusting the amplitude and phase of the sampled receiver outputs in such a way as to minimize the integrated noise power leakage, while preserving the gain/sensitivity of the beam being synthesized.

Figure 20.23 compares conventional processing and SAP applied to the same block of radar data. A reduction of some 20 dB is achieved in this example, where noise, not clutter, is the problem. The pattern here has adapted to minimize the total energy collected in the doppler bands [–5,–1] Hz and [1,5] Hz, whilst maintaining the array

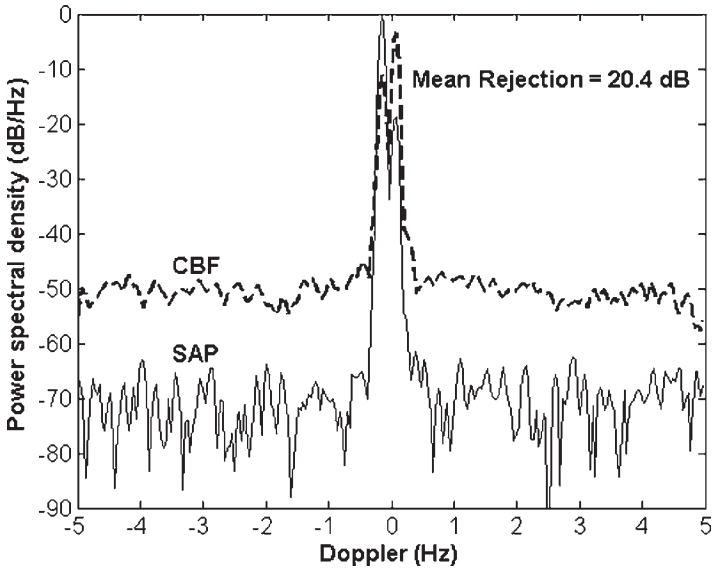


FIGURE 20.23 Comparison of doppler spectra estimated by (i) conventional beamforming (CBF) and (ii) spatially adaptive processing designed to minimize energy outside the low-doppler clutter. The noise (and any fast clutter) has been reduced by ~ 20 dB, which improves detectability of fast targets; note that the sea clutter spectrum changes as a consequence of the new array pattern, i.e., the clutter spectrum does not necessarily represent the clutter spectrum from the cell of interest.

response in the specified direction. Of course, depending on the angular distribution of the noise sources, the array response pattern may have very high sidelobes, but they will lie in directions where the noise levels are lowest. The same may not be true of the clutter that occupies the low doppler band, [–1,1] Hz, which has not been used to guide the adaptation of the array response. Indeed, the resultant clutter spectrum may well have characteristics very different from those found in the steer direction. This is the case in the example shown in Figure 20.23, where the Bragg line ratio has changed dramatically and even reversed in sign.

In the example just given, the SAP weights were kept fixed for the entire CIT. When the external noise field is changing rapidly, as happens when the ionosphere is disturbed, it is necessary to adapt the weights during the CIT to maintain effective rejection of the noise as its azimuthal distribution changes, even for short (~1 s) dwells as used for aircraft detection. This is the province of space-time adaptive processing or STAP.[136] Here, the data used to determine the weights requires not just a single snapshot or average of snapshots but a number of snapshots of the array outputs; this block of data is then used to construct weights that are applied to the block of data before beamforming and doppler analysis. STAP is of particular importance to ship detection, where the external noise field almost invariably changes substantially during the long CIT. The complexity of STAP in this context arises from the fact that each time the weights are changed according to the SAP rules, the main beam experiences a phase shift, even though its amplitude gain/sensitivity is preserved. Thus, over the entire CIT, a sequence of phase shifts is applied, that is, a modulation, which is imposed on the received signal. As a consequence, the strong clutter echoes are spread in doppler, masking targets. To overcome this problem, Abramovich et al.[139] developed a technique known as the *method of stochastic constraints,* which uses different rules for weight adjustment, preserving not only the gain but also, to a good approximation, the phase of the clutter received via the main beam response.

Although these modern spatio-temporal adaptive processing techniques have been proven to be highly effective,[136] the computational and data requirements are so high that most of them cannot be implemented in their most powerful forms for real-time processing. Instead, simplified algorithms with impressive but nonetheless suboptimum performance are employed.[137,138]

In view of the number of simultaneous spatial cells to be considered, detection is generally based on a constant false alarm rate (CFAR) algorithm adapted to the particular clutter environment. In most implementations, the CFAR detection threshold is computed for each resolution cell as a linear combination of order statistics extracted from ranked sample values over a window spanning neighboring range bins, antenna beams, and doppler cells, with provisions for adapting the window shape near strong variations in noise or clutter power. Often the results are similar to those predicted by the log-normal distribution, as evidenced by the experimental data shown in Figure 20.22*b*, which gives the power-level distributions for the example shown in Figure 20.21. These approximately log-normal distributions are typical for benign conditions.

Tracking. Perhaps the most fundamental difference between skywave radar and other radars is the existence of multiple propagation paths, with distinct time delays, angles of incidence, doppler shifts, and fluctuation properties. The tracking stage must deal with the multiplicity of echoes associated with each individual target and, by extracting and assimilating information about the prevailing ionosphere, infer the number of genuine targets, their true locations and velocities, and perhaps

other information such as target altitude.[140] (Estimates of aircraft target altitude are very useful, but skywave radar has not proved to be a reliable means of obtaining accurate estimates.)

The problem of converting from radar coordinates to geographical coordinates is referred to as *coordinate registration (CR).* Dozens of CR techniques have been explored, including (i) inference from a regional ionospheric model, (ii) deploying a network of repeaters or beacons in the radar footprint, (iii) correlating coastlines with land clutter—sea clutter boundaries in the radar data, (iv) correlating other parameters such as scattering coefficient, (v) using known target information such as reports from ships and commercial airline flights, and (vi) registering airports, where tracks originate or terminate. The key to robust CR is the fusion of all available information in a consistent probabilistic framework.[141] Most radars implement target tracking as a separate stage that operates on the candidate targets after they have been registered, initiating, updating, or terminating tracks, as appropriate. Experience with many different types of tracking schemes has led a number of operational radars to converge on algorithms based on variants of probabilistic data association (PDA),[142] sometimes generalized to maintain multihypothesis models.[143] Unlike traditional tracking filters such as the Kalman filter, which selects a single detection (i.e., peak or plot) to associate with each maintained track, PDA filters combine the influence of all the candidate peaks within a prescribed radius to compute a track update. In the skywave radar context, this has yielded superior results.

An important decision relates to where the coordinate registration is implemented. Some systems establish tracks in radar coordinates and then pass the tracks, including multiple tracks from a single target, to the CR system, which must identify and reconcile any multiple tracks as well as perform the registration. Alternatively, the problem of target tracking can be integrated with the problem of determining the ionospheric propagation paths. By augmenting the state vector used for representing a target with additional parameters that characterize the propagation path structure, a joint estimation problem can be formulated and solved.[144–147] In this way, the targets contribute to registering their own coordinates.

20.12 RADAR RESOURCE MANAGEMENT

The complexity of the HF radar operating environment—the ionosphere, clutter, noise, and other users of the band—necessitates a commensurate complexity in system design so that the radar can adapt to prevailing conditions by selecting the best frequency (or combination of frequencies[148]), waveform, signal processing, detection thresholds, and so on, for the task at hand. Achieving this optimum control is important because experience has shown that HF skywave radar performance may degrade dramatically with only modest departures from optimum settings. Thus, there is a demand for two ingredients: (i) information about the environment and (ii) a mechanism, or at least a strategy, for using that information to control the radar parameters.

At the very minimum, HF skywave radars must maintain a real-time awareness of propagation conditions as a function of frequency, range, and bearing (azimuth), as well as a detailed knowledge of spectrum occupancy. This is usually achieved by providing auxiliary facilities, including some or all of the following: (i) Conventional ionospheric sounders (vertical and oblique incidence), which determine the ionospheric electron density profile by measuring the time of flight of reflected radiowaves over a range of frequencies; this electron density profile information is assimilated into

local real-time ionospheric models. (ii) A wideband backscatter sounder, that is, a low-power, low-resolution radar that sweeps across the HF band, measuring echo strength versus time delay (group range), to see which frequencies are illuminating any given region. (iii) A mini-radar, similar to a backscatter sounder, that employs a narrowband waveform to study the doppler structure of the echoes as a function of group range for selected frequencies. (iv) A network of remote beacons or transponders to provide coordinate registration. (v) Spectrum monitoring receivers to locate clear channels for possible use and to assess their properties. A detailed description of the suite of auxiliaries for the Jindalee radar can be found in Earl and Ward.[149,150]

Traditionally, the control mechanism has been the expert radar operator. For many reasons, this is not wholly satisfactory, so various alternatives have been explored, including packaged "recipes" that can be invoked by less skilled operators and expert systems implemented with artificial intelligence constructs.

Apart from the need to adapt to the changing ionosphere and noise environment, HF radars are frequently tasked with a variety of missions to be conducted more or less concurrently, with time-varying priorities. These will generally involve different waveforms, task-specific constraints on frequency (in addition to propagation considerations), differing requirements in terms of acceptable propagation quality, and so on. A common example is the desire to search for ships and aircraft concurrently. Accordingly, optimization of the allocation of resources becomes a critical issue, with significant implications for the way skywave radars operate.

One approach to fitting more tasks into the time available is to partition the radar transmit and receive arrays, together with the transmitter modules and receivers, so that when conditions are favorable, the partitions can operate as independent radars with reduced sensitivity and resolution. For example, the Jindalee and JORN radars are dynamically reconfigurable as full or half radars. Alternatively, it may be acceptable to employ a dual-purpose waveform, able to support two distinct missions, though with suboptimal performance, or for the transmit system to radiate multiple orthogonal waveforms for simultaneous reception and processing at a proportional loss of average power on each transmission but without loss of spatial directivity. But even if these approaches are sometimes applicable, it is almost always necessary to schedule the various tasks such that acceptable revisit times are maintained on important missions, with low priority tasks, such as weather monitoring, activated less frequently, and challenging tasks activated when suitable propagation conditions occur.[151]

One other important issue arises with resource management and that is the ability to diagnose faults and failures as soon as they occur. Not only is this necessary to hasten the repair procedure, but also it enables the radar to adapt its configuration to do the best it can with impaired equipment. For example, if the receiver connected to the middle element in an antenna array were to fail, beamforming would be more seriously degraded than if the failure were to occur in a receiver connected to an element at the end of the array. Automatic detection of such a failure and reallocation of an end element receiver to the middle element of the array would minimize the degradation.

20.13 RADAR PERFORMANCE MODELING

Modeling is an essential part of the radar design process and also as a means to predict the performance of existing or proposed radars to which one does not expect access. In these roles, the emphasis is on fidelity. In HF skywave radar systems, modeling has

another role—as a real-time monitor of what is expected under the prevailing conditions, so as to alert the radar operator if a discrepancy emerges, which might be indicative of natural, equipment-related, or intentional events warranting attention.

An issue one encounters when discussing radar performance is the choice of performance criterion. From the user's perspective, it may seem logical to measure performance in terms of time taken to establish a track on a given target, averaged over time and coverage, since tracks are the essential product delivered to users by the radar. Of course, this places the onus on the tracking system, so one might step back a little and choose instead SNR achievable on a given target, again averaged over time and coverage. But this ignores the precision of the measurement, associated with both resolution, which may impact on the utility of the detections, and accuracy, in the sense of registration in absolute (geographical) coordinates. And what of the smallest detectable target? Coverage and coverage rate? Ship detection capability versus aircraft detection capability? Clearly, there is no single preferred metric.

The measure adopted here for illustrative purposes is the achievable signal-to-noise ratio (SNR) as defined by the radar equation (Eq. 20.2). The constituent variables that appear in the equation have been discussed in Sections 20.4–20.11, emphasizing the unique considerations that arise with HF skywave radar, so the tools for performance modeling and analysis are available. By considering the case of noise-limited detection, as applies especially to aircraft targets, the additional complications of clutter-limited detection are avoided. In the following paragraphs, two different approaches to performance modeling for the case of noise-limited detection are described in some detail.

The NRL-ITS Radar Performance Model. The NRL-ITS Radar Performance Model developed by Lucas et al.[152] provides a number of tools for radar performance estimation. It does not employ full 3D ray-tracing, such as the code described by Jones and Stephenson,[50] which can provide paths in three dimensions, including delays and losses for both ordinary and extraordinary rays (see Section 20.4). Arguing that such comprehensive calculations are excessive when there are significant uncertainties in the electron density distribution, NRL Memorandum Report 2500[152] describes the basic technique used for path determinations. A simple closed-form virtual path trace, Snell's law for a spherically symmetric medium, is sequenced through elevation radiation angles in 1° increments. This process is incremented in 1-MHz steps over the radar's operating band. In the present example, a vertical sounding of the ionosphere 700 km downrange has been used as the electron distribution for all one-hop paths, and a sounding 1400 km downrange is used for two-hop paths. Figure 20.24 gives night and day examples of ionograms showing the ionosphere 700 km downrange from a radar located at 38.65°N and 76.50°W looking east, in summer at a median sunspot number of 50. The corresponding winter case is presented in Figure 20.25. Such plots can be used to determine the maximum frequency that will propagate to a given range; in most cases, the optimum frequency is just below the maximum frequency. For the little tables shown in Figure. 20.24 and 20.25, FC is the critical frequency, in megahertz; HC is the height of maximum ionization or the nose of the parabola, in kilometers; and YM is the semilayer thickness, in kilometers. Es describes the statistical range of variation of sporadic-E as M (median), L (lower decile), and U (upper decile) critical frequencies, in megahertz.

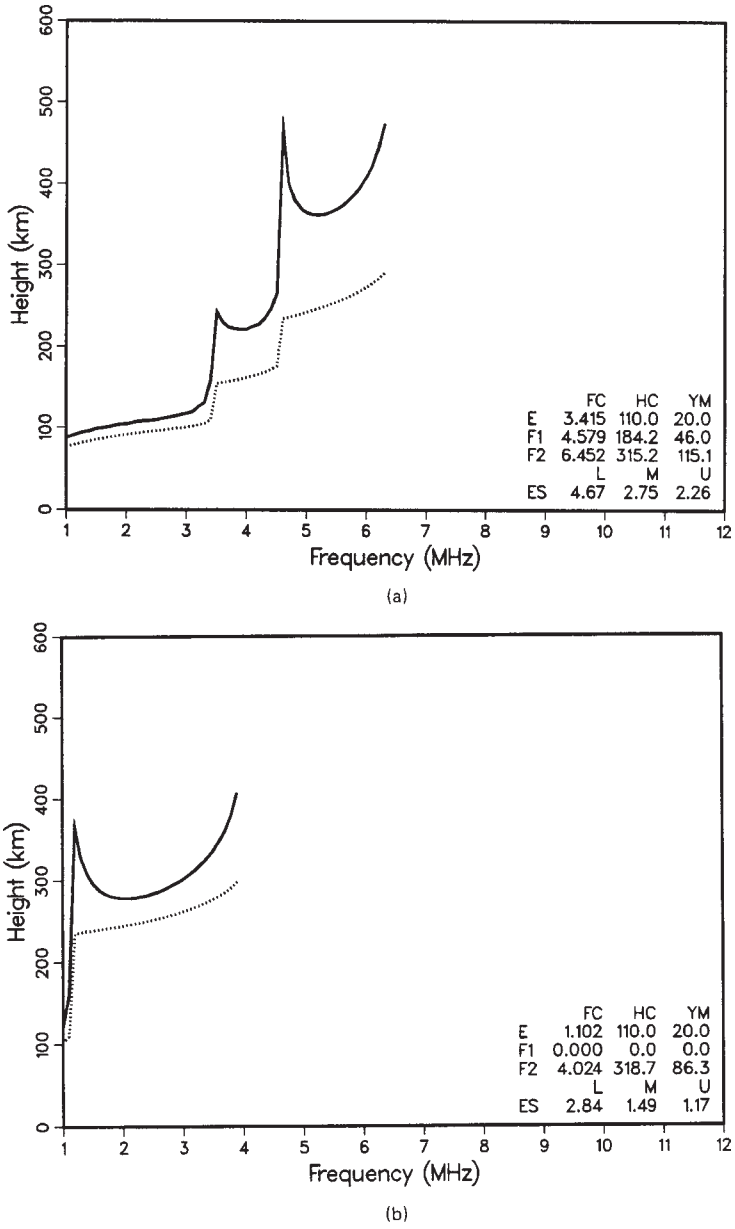
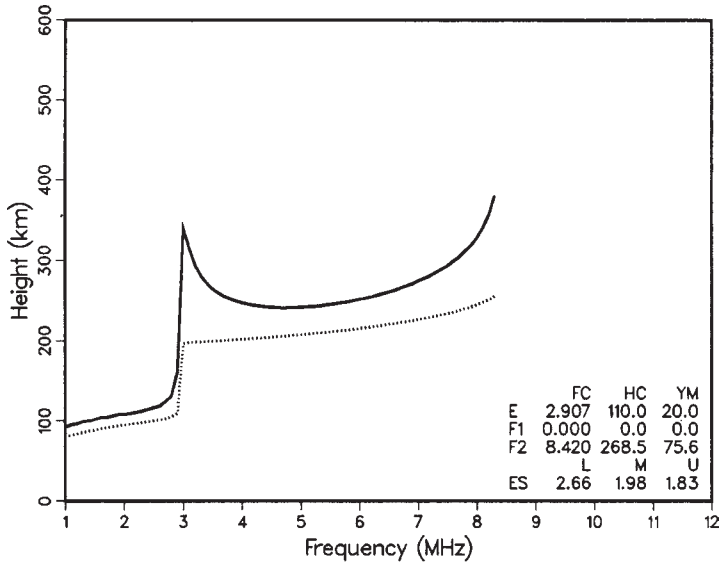
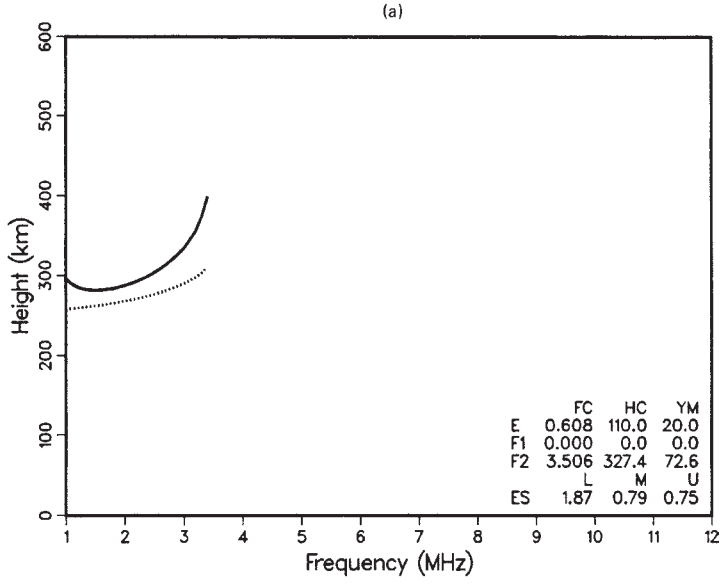


FIGURE 20.24 The virtual (solid line) and true (dashed line) reflection heights are given for July, SSN = 50, and a mid-Atlantic coast refraction area: (*a*) 1800 UT is a daytime example, single hop and (*b*) 0800 UT is a nighttime example. See text for an explanation of the table in the lower right of the figure.



(a)



(b)

FIGURE 20.25 Predicted ionograms as in Figure 20.24, but for January: (*a*) 1800 UTC for day and (*b*) 0800 UTC for night. See text for an explanation of the table in the lower right of the figure.

Constant plasma-frequency contours versus range from the radar for 0800 UTC, SSN 50, January (night) and 1800 UTC (day) are presented in Figures 20.26 and 20.27 to illustrate the tilt of the ionosphere. For the night case, the concentric

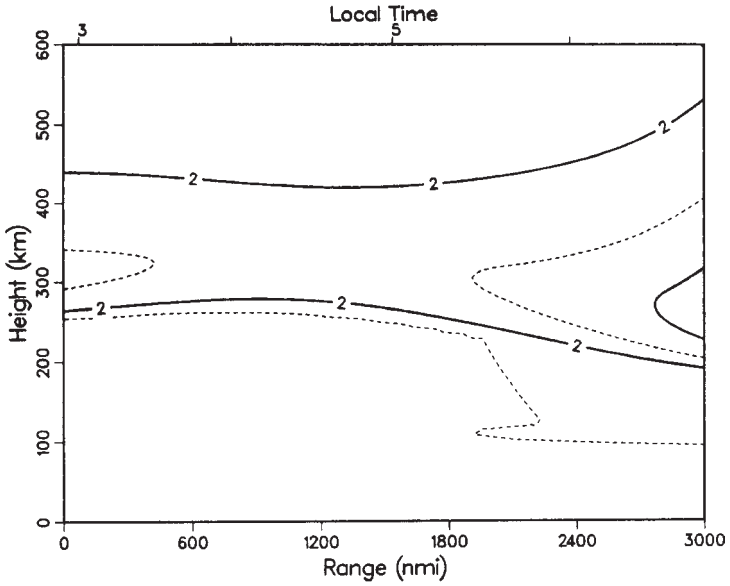


FIGURE 20.26 Plasma-frequency contours are given in MHz, extending from the radar to a point east 3000 nmi downrange, for a January night example.

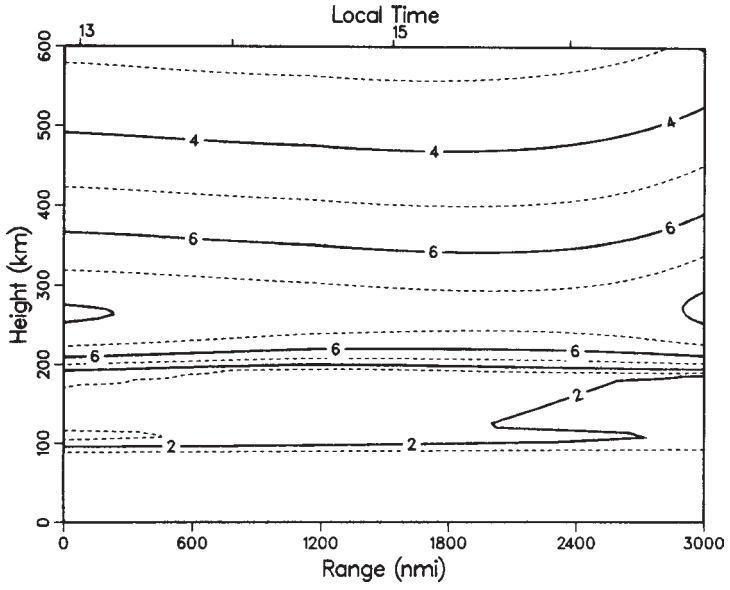


FIGURE 20.27 Plasma-frequency contours are given in MHz, extending from the radar to a point east 3000 nmi downrange, for a January day example.

spherical assumption from the 700-km downrange position will give paths that are slightly long for one-hop ranges. In the two-hop ranges, the no-gradient assumption causes more distortion. In general, errors of this nature have little impact on performance prediction. However, near-real-time analysis for virtual range and azimuth correction to great-circle distance and bearing (grid registration) requires that tilt or gradient effects be taken into account. The daytime example has little horizontal gradient, and the simplifying assumption makes little difference. When better accuracy is desired, the correct vertical profile can be used for each radiation angle; also, gradients can be simulated by making the ionosphere nonconcentric with the Earth. A more complete path analysis should be used in radar performance assessment and management, but these plasma density contours can be used to estimate the magnitude of the errors introduced by the assumption of a spherically symmetric ionosphere.

Figure 20.28 shows a performance prediction for a hypothetical radar in the form of an oblique sounding. A typical skywave radar is supported by one or more vertical sounders and oblique backscatter sounders for transmission-path analysis and to aid in radar-frequency management. Of course, the radar itself is an oblique sounder, but its sounding data is restricted to the frequency, waveform, and scan program of its primary surveillance task. An adjunct oblique sounder can provide information in

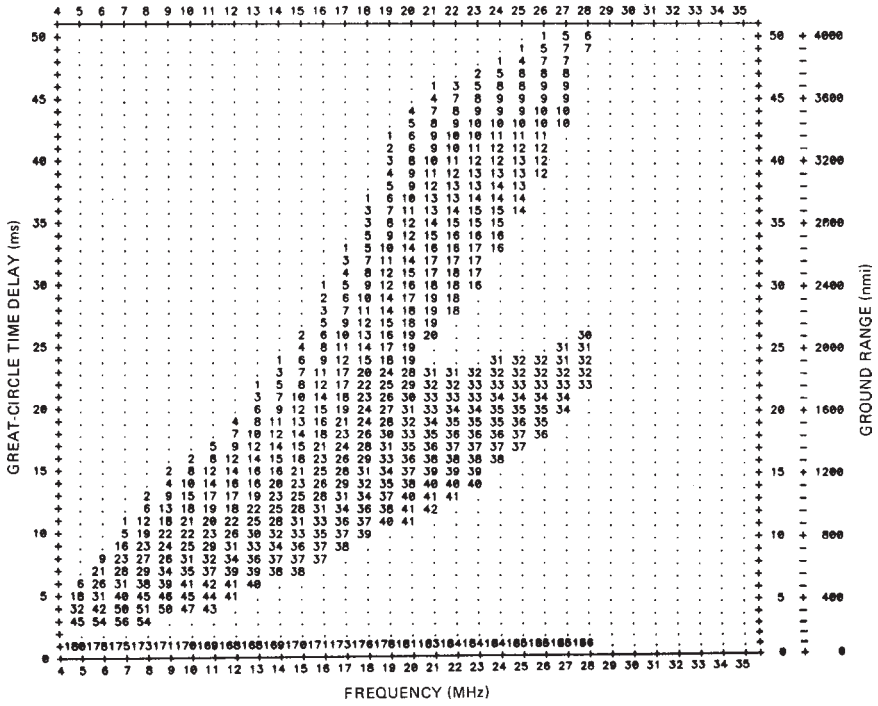


FIGURE 20.28 The numbers in this figure show the SNR in dB as a function of frequency and range in the form of a typical oblique backscatter sounding: January, 1800 GMT (day), SSN 50, location 38.65°N and 76.53°W, bearing 90°. See the text for a description of radar parameters.

the form of Figure 20.28 on Earth backscatter echo power. Estimates for noise power spectral density are derived from CCIR Report 322-3, as described in Section 20.9. In this prediction, SNR in decibels is plotted as a function of operating frequency and great-circle time delay or ground range. The numbers just above the abscissa (at 1-ms delay) are the noise powers in decibels below 1 W/Hz. For this plot, the UTC time is 1800, SSN = 50, $P_{av} = 200$ kW, $G_t G_r = 50$ dB, $T = 1$ s, and $\sigma = 20$ dBsm. Figure 20.29 gives the corresponding night plot.

The shape of these displays is quite similar to what would be seen with a diagnostic oblique sounding; the levels would generally be greater because the resolution cell area times the surface scattering coefficient is generally much larger than 20 dBsm. Some of the night-day contrasts, such as available frequencies and difference in noise level for the same range, are evident. Also note that at night the 5-MHz lower frequency limit does not provide coverage closer than about 900 km. It should be remembered that this is a median SSN 50 calculation, and if consistent performance for ranges as close as 900 km is required during nights, a lower frequency limit should be selected to deal with periods of lower solar activity and the critical frequency distribution. The plots show that operation on a single frequency may experience less than ±3 dB variation over a range interval of 1000 km. Also, if frequency selection had been made with a 2-MHz granularity instead of the 1 MHz used, the SNR would be reduced by only a decibel or so.

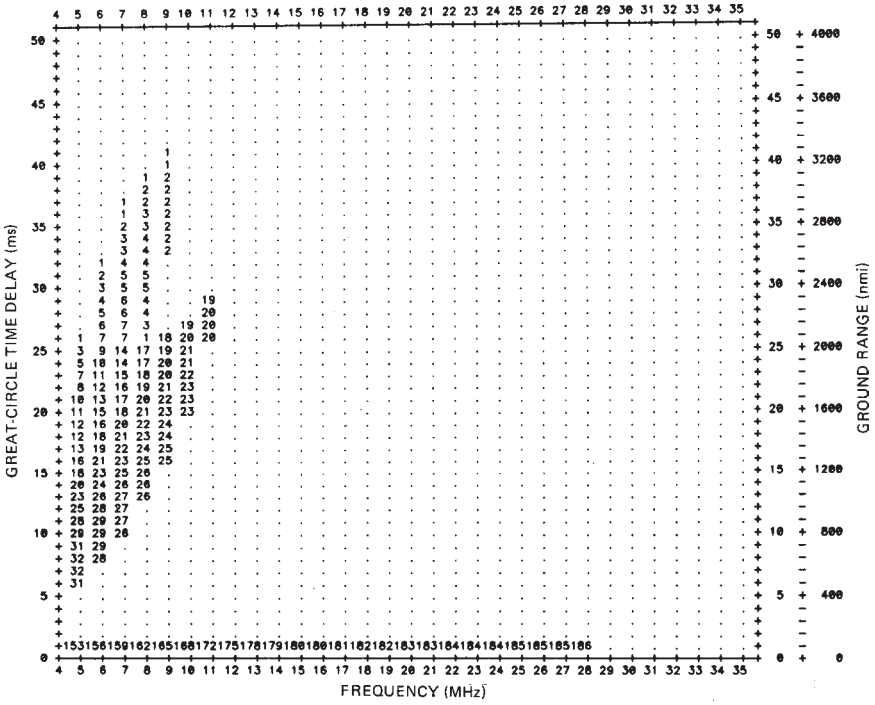


FIGURE 20.29 The numbers in this figure show the SNR in dB as a function of frequency and range in the form of an oblique sounding, as in Figure 20.28, but for 0800 GMT (night). See the text for a description of radar parameters.

The performance estimation charts that follow come from analyses as described above. After performing oblique sounding calculations, a range-indexed table of propagation and noise parameters is compiled. Parameter selections are made on the basis of the best SNR in each nominal 50-nmi interval, but the selection is adjusted to come from the adjacent lower frequency to avoid an optimistic bias. Then parameter plots are generated as a function of range. The variables adopted here as parameters are propagation losses, frequency, noise, and elevation radiation angle. The choice of range as the independent variable may seem artificial, but it is a useful approach for performance examination. With these curves, the impact on radar SNR performance can be estimated for selected antenna gain patterns, transmitted powers, target RCS, and coherent integration times (CIT).

Figure 20.30 is an example for January with moderate solar activity, SSN = 50, in daytime. The R^4 loss is the fourth power of range to the target in meters, expressed in decibels. $R^4 + L$ adds nondeviative absorption, deviative absorption, sporadic-E obscuration, and ground-reflection losses if there is more than one hop. The sharp increase in loss just before 2000 nmi is caused by transition from one to two hops; for two hops, the lossy D region is transited twice as many times, ground-reflection loss is added, and required operation at a lower frequency increases loss. The jagged curve in the transition region is due to the parameter selection process; in radar operation, the frequency would be selected to minimize transition effects. The frequency, radiation angle, and noise power per hertz that go with this site and look direction are also plotted.

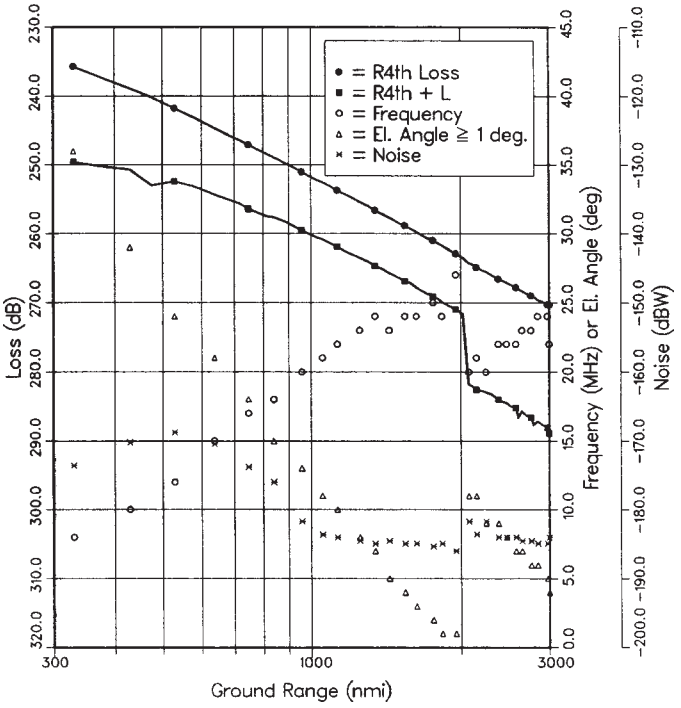


FIGURE 20.30 Radar performance-controlling variables plotted as a function of range for January, 1800 UTC, SSN 50.

An example will serve to illustrate the use of such plots. Select 1000 nmi as the range. Then, the frequency is 20.5 MHz (wavelength = 14.6 m and $\lambda^2 = 23$ dB), noise power = –182 dB, and $R^4 + L = 260$ dB. Choose 53 dBW for P_{av}, 20 dB for G_t, 30 dB for G_r, 0 dBs for T, 20 dBsm for RCS, and 6 dB for F_p. Substituting in Eq. 20.2,

$$SNR = 53+20+30+23+0+20+6-33-260-(-182)=41\ dB$$

Figure 20.31 shows the performance indicated with these assumptions for all ranges. A path factor enhancement of 6 dB has been chosen as a representative value of constructive multipath interference for an aircraft target. The beamwidth has been taken to be 5.7° and the surface scattering coefficient to be –35 dB, and with a 12 dB path enhancement, the clutter level has been plotted. The clutter-to-noise ratio (CNR) at 1000 nmi is about 88 dB. For the constant beamwidth assumed, the clutter-to-signal ratio increases with range and is 47 dB at 1000 nmi. Large clutter-to-signal ratios are typical of HF radar; some form of doppler filtering is used to separate targets from clutter.

In Figures 20.32, 20.33, 20.34, and 20.35, performance estimation curves are given for winter and summer seasons, night and day, and low solar activity. The permissible frequency selection is set between 5 and 28 MHz, and antenna radiation is not considered below an elevation angle of 1°. The analyses were made for a radar off the mid-Atlantic coast of the United States and should be a good approximation for any location where transmission paths are through the middle magnetic latitudes.

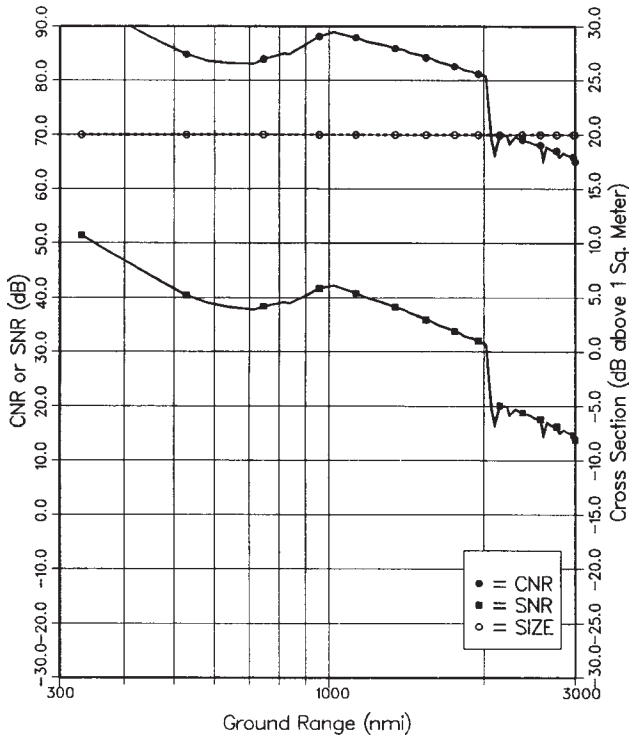


FIGURE 20.31 A specific example of SNR and clutter-to-noise ratio (CNR) as determined by using Figure 20.30. The target RCS is labeled "size" and is treated as constant.

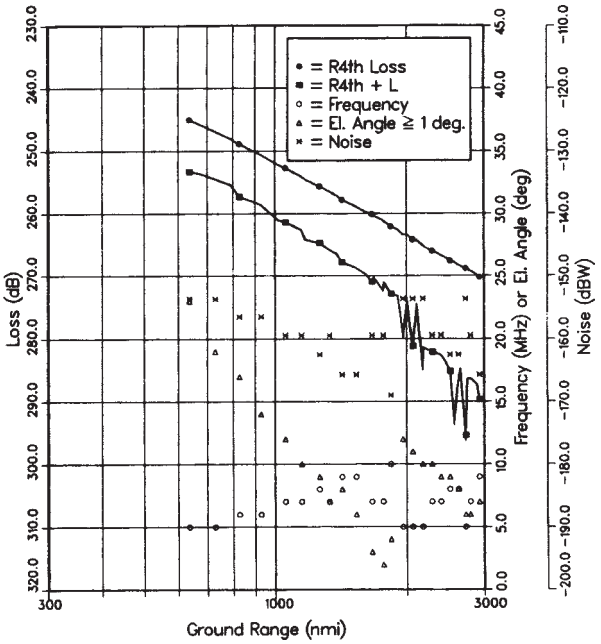


FIGURE 20.32 Radar performance estimate for January 0800 UTC, SSN = 10

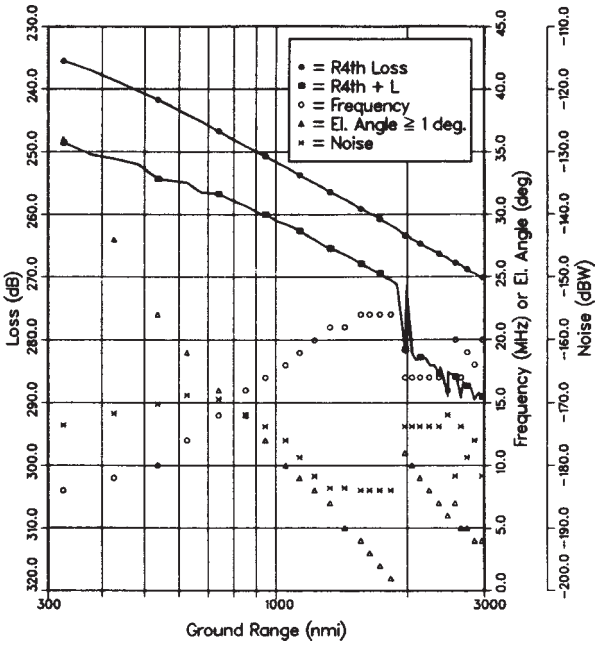


FIGURE 20.33 Radar performance estimate for January 1800 UTC, SSN = 10

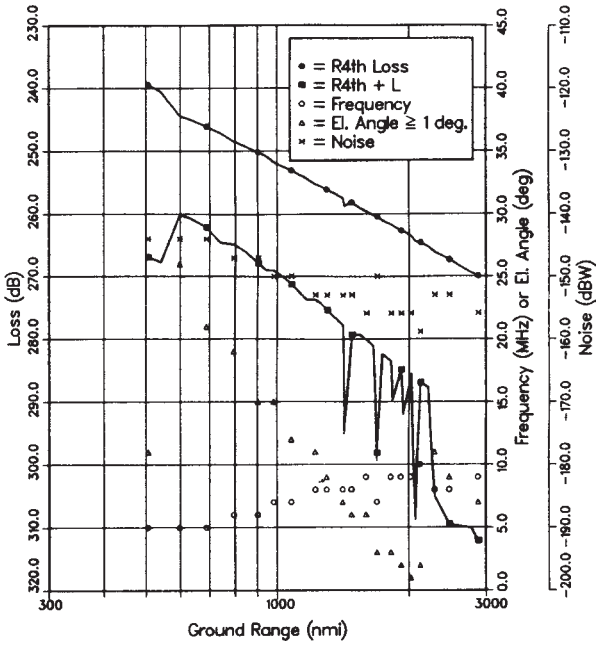


FIGURE 20.34 Radar performance estimate for July 0800 UTC, SSN =10

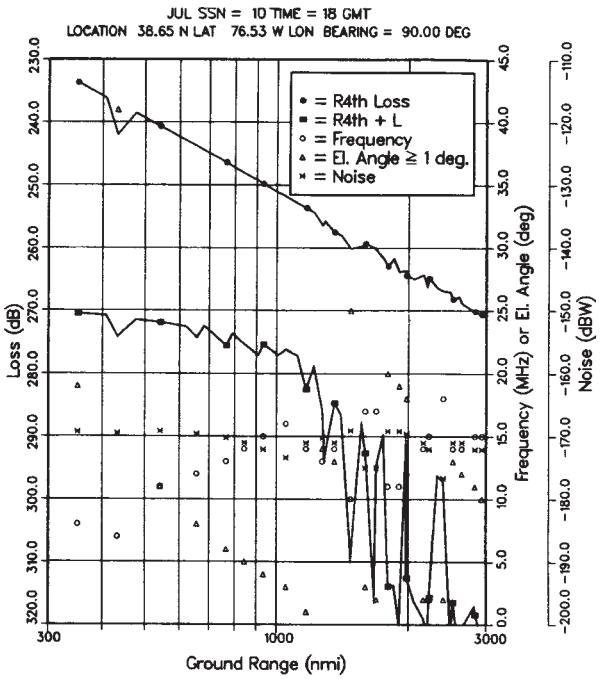

JUL SSN = 10 TIME = 18 GMT
LOCATION 38.65 N LAT 76.53 W LON BEARING = 90.00 DEG
• = R4th Loss
■ = R4th + L
○ = Frequency
▲ = El. Angle ≧ 1 deg.
× = Noise
Loss (dB)
Frequency (MHz) or El. Angle (deg)
Noise (dBW)
Ground Range (nmi)
300
1000
3000

FIGURE 20.35 Radar performance estimate for July 1800 UTC, SSN = 10

These performance estimation curves can be used to show extremes in SNR variation, antenna elevation angles required, and expected clutter-to-noise ratios. The expected performance of a particular radar design can be explored with these graphs because summer and winter do give good coverage of the important variables. The performance curves have been limited in almost all examples to the case of low solar activity, since, in general, this is the most difficult time. The higher frequency availability and performance afforded at high solar activity is illustrated in Figure 20.36, which treats the case of July 1800 UTC for SSN = 100.

Analysis of performance estimation curves for all combinations of (i) the four seasons, (ii) day and night, and (iii) high and low solar activity[153] reveals consistent behavior:

(i) Summer shows much greater losses than winter.

(ii) Except for summer, night losses are only slightly less than day losses.

(iii) Night noise is much greater than day noise.

(iv) For a specific range, optimum frequencies vary by 3:1.

This OTH performance presentation format can be used to decide on the antenna patterns and powers required for specific targets and missions, or it can be used to exhibit periods of enhanced or degraded performance for an existing design.[154]

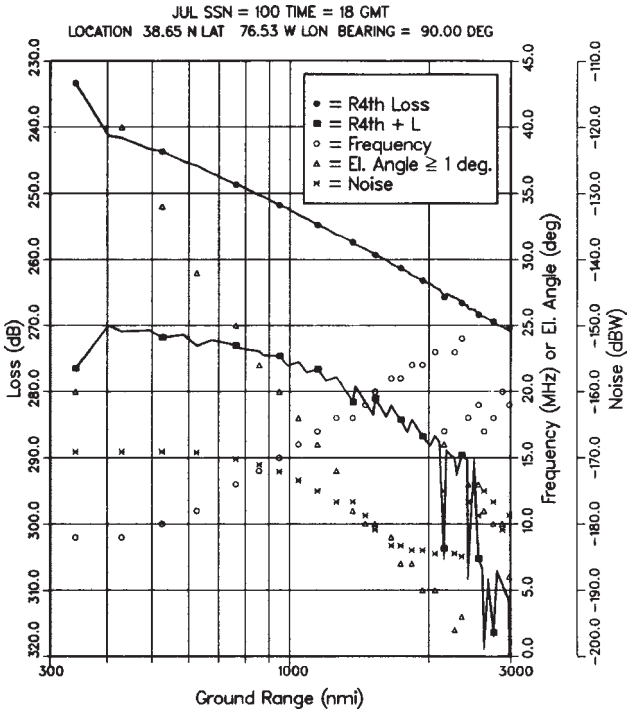


FIGURE 20.36 Radar performance estimate for July 1800 UTC, SSN = 100

Several qualifiers should be kept in mind. At other geographic locations, the appropriate CCIR noise should be selected or, better yet, noise measurements made in situ. For radars that use auroral zone paths, specific analyses are required and target obscuration by spread-in-doppler clutter must be considered. The performance estimates shown in the figures assume that the radar design and waveforms are such that external noise is the control. The use of a single description for night and day gives a fair representation, but the transition from night to day is very abrupt and requires careful frequency management in radar operation. The ionospheric description that has been used is for what has been termed the *quiet ionosphere,* conditions that apply most of the time. Under disturbed conditions, performance may be markedly inferior to that predicted.

The Jindalee Radar Performance Model. The Jindalee Radar Performance Model draws on several unique databases. Since 1984, oblique backscatter soundings have been recorded on a 10 minute cycle, scanning from 5 to 30 or (optionally) 45 MHz.[149,150] Eight simultaneous beams spanning a 90° arc are formed. Prior to 1991, the system used a single receiver scanning the 8 receive beams, servicing each receive beam for 25 kHz of each 200 kHz portion of a backscatter ionogram. Post-1991, each beam has been serviced continuously with its own receiver. Background noise data used in conjunction with the backscatter data for the Excess Power analysis is collected using the same eight directional beams as the backscatter sounder.

While there are a few significant gaps in these time series, they span two solar cycles; further, integrity has been maintained by extensive vetting before entering new data into the database. The unique advantage of this database is that the noise and propagation data are recorded under identical ionospheric conditions, whereas combining independent clutter and noise statistical models such as IRI-2001 with CCIR 322-3 preserves no correlations, however strong they may be.

Several forms of data analysis and display are accessible, but perhaps the most useful are the maps of (i) Excess Power and (ii) Optimum Frequency. The Excess Power parameter is constructed as follows. Consider a specific target whose RCS is known or estimated as a function of frequency. Suppose a prediction of median radar detection performance is required for a specific month of the year and a particular sunspot number or year of solar cycle.

Analysis is initiated by selecting from the database a month that matches the requirements. Individual backscatter ionograms are paired with the concurrent background noise data and the ratio taken to yield a population of estimates of true subclutter visibility (SCV) for each frequency step, each beam, and each range bin. Next, using a specified time granularity—typically one hour—median values of SCV are calculated for each spatial cell and time interval from the statistical population of individual SCV estimates. Note again that these median values are the median values of the SCV population, not a statistic derived by combining median values of clutter with median values of noise.

The median SCV values are related directly to the backscatter sounder transmit power and the transmit antenna gain. From the radar equation (Eq. 20.2), the SCV can be scaled to that which would be expected for the main radar, with its different radiated power P_T, aperture, radar bandwidth, etc. Combining these scaled SCV values with (i) a model of the surface backscattering coefficient versus frequency (almost all the OTH radar coverage is ocean, so based on regional wave statistics, the constant value –23 dB is used); (ii) the target RCS versus frequency model; and (iii) an estimate of the target echo signal processing loss, typically ~ 12 dB, which arises predominantly from FFT analysis, yields the predicted median target SNR at each spatial location as

a function of frequency. This can be used directly, but it is convenient to define the Excess Power at each frequency as follows.

Suppose:

- An SNR threshold of M dB is required before the signal processing will register a detection and
- The SNR calculated for a particular range-azimuth cell in the coverage is SNR(f) dB (which may be negative)

Then, in dB,

$$\text{Excess Power } (f) = M - \text{SNR}(f) + P_T - P_{\text{REF}}$$

is the amount of extra power, relative to any chosen reference power P_{REF}, needed to achieve detection as a function of frequency. A negative value indicates that the radar has greater sensitivity than required. Figure 20.37 shows an example presented as a contour plot. The second parameter of special interest is optimum frequency, defined as the value of the frequency at each range-time location that maximizes the SNR for the given target, taking all radar equation factors into account. Again, a contour plot format is used here in Figure 20.38.

Other Modeling Approaches. The models described above are formulated in the context of the radar equation (Eq. 20.2). An alternative is to employ a coherent "process model," in which the field amplitude and phase are tracked from transmitter to receiver. This approach has been used to simulate the effects of multipath, diffuse scattering, polarization effects, and nonlinearity.[69,70,155] For example, Figure 20.39 shows

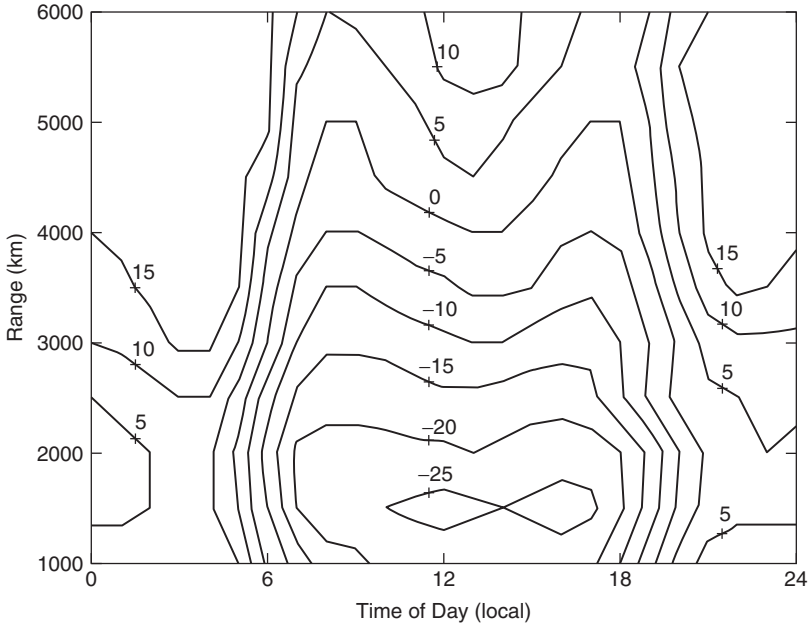


FIGURE 20.37 Range-time map of power required, relative to a P_{REF} of 100 kW, to detect a specific target, as a function of range and time-of-day, for a particular month and level of solar activity

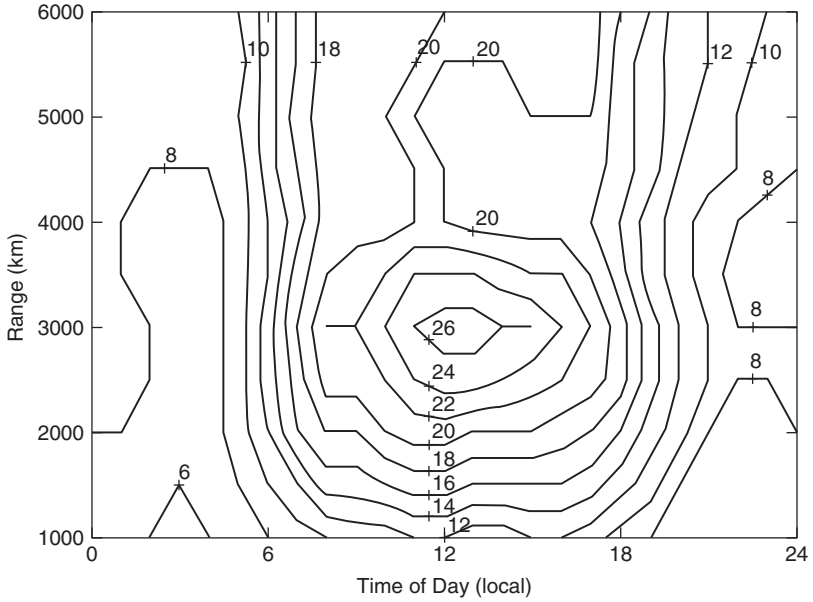


FIGURE 20.38 Range-time map of optimum frequency in MHz to detect a specific target, as a function of range and time-of-day, for a particular month and level of solar activity

the predicted distribution of the resultant radar cross section when multipath and Faraday rotation are taken into account. The relevant physics—rough surface forward scattering coefficient, bistatic (in the vertical plane) free-space target scattering cross

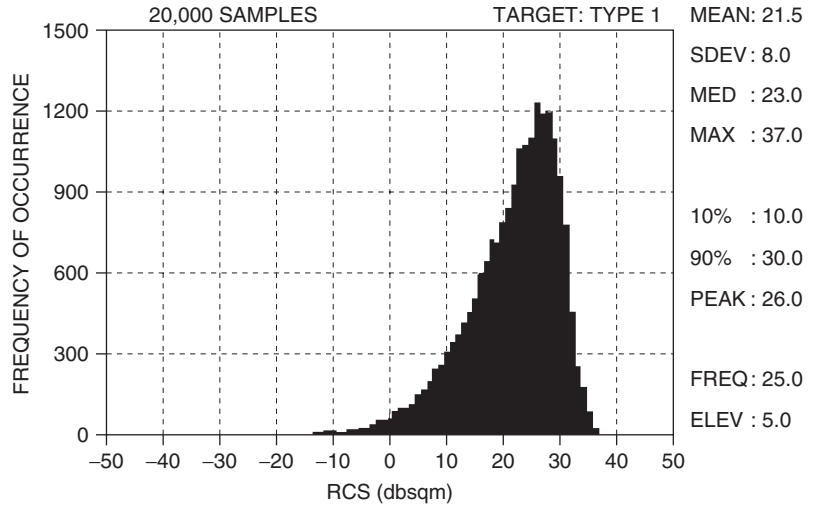


FIGURE 20.39 Predicted distribution of effective RCS for the Aermacchi MB326H trainer jet aircraft viewed nose-on at a height of 1000 feet when ground-reflection multipath and Faraday rotation are taken into account.

section for hybrid multipath, Faraday rotation, and differential Faraday rotation—are incorporated via parametric models derived from measurements or computational electromagnetics and the distribution obtained by Monte Carlo simulation.

APPENDIX: HF SURFACE WAVE RADAR

General Characteristics and Capabilities. Although skywave propagation provides the unique capability of low altitude target detection at ranges of thousands of kilometers, other forms of propagation at HF can be exploited in radar applications. By far the most common of these is *ground wave* or *surface wave* propagation, which is most effective for vertically polarized radiowaves traveling over highly conductive surfaces such as seawater. In addition, there are applications for which line-of-sight or spacewave propagation is appropriate, such as measurement of the HF RCS of aerospace vehicles. Furthermore, in many instances, bistatic configurations can be employed, with the possibility of using different propagation mechanisms for transmitter-target and target-receiver paths. In view of the familiarity of space wave propagation and the preceding discussion of skywave systems, it suffices here to address the main features of HF surface wave (or ground wave) radar (HFSWR).

HFSWR systems tend to fall into two categories: (i) low-power radars intended primarily for oceanographic remote sensing, especially of ocean currents, and (ii) larger and more powerful systems with target detection as their primary mission. The former are in widespread operation around the world; only a few of the latter are operational in surveillance roles. The ability of the low-power remote sensing radars to detect ships at over-the-horizon ranges, although modest, has been exploited in some dual-purpose deployments.

The principal virtue of HFSWR as an ocean surveillance radar lies in its ability to detect small surface vessels and low-flying aircraft at ranges far beyond the visible horizon. As with skywave radar, performance depends strongly on environmental and target parameters, as well as radar design; the detection ranges cited in Table 20.3 provide some indication of capability against surface vessels and low-flying aircraft, as claimed or reported for several established HFSWR systems.

TABLE 20.3 Nominal or Claimed Maximum Detection Ranges (kilometers) of Some HFSWR Systems
(SeaSonde is a compact low-power radar designed primarily for remote sensing; it is available with an upgraded performance option. The other radars shown were designed for surveillance. In all cases, performance may fall far short of the cited values under inclement environmental conditions.)

TARGET TYPE	**SWR 503** Raytheon Canada	**HFSWR** BAE Systems UK	**SECAR** Daronmont Australia	**Podsolnukh-E** Niidar Russia	**SeaSonde** Codar US
Frigate	520	> 200	> 400	260	190
Offshore trawler	450	85	> 250	180	120
Small fishing boat			120		65
Go-fast speed boat			70		35
Rigid inflatable boat		22			
Low-flying fighter-sized aircraft		75	> 200		

Much of the discussion relating to skywave radar carries over directly to HFSWR, but there are a few areas where differences are significant:

- The antennas must be designed and positioned to achieve high efficiency in coupling to the surface wave mode. Experiments have shown that higher field strengths are generated at over-the-horizon ranges when the transmitting antenna is located close to sea level, rather than in an elevated position; placing the antenna even just one or two wavelengths above sea level can introduce several dB of additional loss.[156]
- Most HFSWR systems use a very broad "floodlight" transmit beam to illuminate the entire coverage arc; multiple simultaneous receive beams are formed to fill the arc and update all tracks simultaneously. This reduces cost and complexity but incurs some loss of sensitivity for noise-limited target detection.
- Coherent integration times may extend to hundreds of seconds, as HFSWR is not reliant on the ionosphere as a medium for propagation.
- Only vertically polarized electromagnetic waves propagate effectively in the surface wave mode over the sea, so the HFSWR signal-receiving antennas are necessarily vertically polarized. On the other hand, unwanted signals and interference arriving by skywave may have any polarization. This presents a means for rejection of interference by filtering in polarization space: any signals correlated with those received on an auxiliary horizontally polarized receive antenna can be cancelled from the outputs of the vertically polarized array.
- The increasingly rapid attenuation of propagation to ranges beyond about 200 km, as shown in Figure 20.40, means that at longer ranges, relatively little detection performance improvement is gained by large increases in transmitted power.
- The rate of decay of the surface waves rises sharply with increasing frequency, as shown in Figure 20.40, whereas the RCS of small targets tends to increase rapidly and external noise decreases. It follows that radar design is sensitive to the classes of targets to be detected.
- While HFSWR does not rely on the ionosphere, echoes from irregularities in the ionosphere may appear at ranges in excess of ~100 km, Similarly, ground reflections received via oblique skywave propagation may appear at ranges in excess of ~200 km. Such echoes may be spread in doppler and can constitute a serious problem for HFSWR systems. Accordingly, antennas should be designed with low gain at higher elevation angles.

The main design parameters of a representative HFSWR system—the Daronmont SECAR radar—are listed in Table 20.4. (The corresponding values for skywave radars can be found in Table 20.2.)

Propagation Considerations in HFSWR Systems. To a good approximation, when a target of interest is above the optical horizon of an HFSWR, the field incident on the target can be decomposed into terms that correspond to (i) direct line-of-sight, (ii) sea-surface reflection, and (iii) a lateral or "surface-attached" wave. Beyond the horizon, the surface wave is the dominant contributor, but at shorter ranges, all three mechanisms must be taken into account. Accordingly, the relationship between (i) target echo strength and (ii) the range and altitude of the target is not a simple one. Moreover, the calculation of the field distribution is computationally expensive if

TABLE 20.4 Specifications of SECAR, an HF Surface Wave Radar Designed for Surveillance of the 200 nmi Exclusive Economic Zone

Radar	SECAR
Manufacturer	Daronmont Technologies
Type	Bistatic HF surface wave radar
Tx-Rx site separation (km)	50–150
Power (average) (kW)	5
Power (peak) (kW)	5
Frequency band (MHz)	4–16
Waveform	linear FM-CW
Bandwidth (kHz)	10–50
Waveform repetition frequency (Hz)	4–50
Tx antenna design	Single vertical log-periodic antenna with ground screen
Rx antenna design	16 or 32 endfire monopole doublets with ground screen
Rx aperture (m)	200 – 500
Beamwidth 4 MHz	> 9°
16 MHz	> 2°
No. of simultaneous beams	16 or 32
Instantaneous range depth (km)	100–500
No. of range bins	10–200
Coherent integration time (CIT) (s)	1–120
No. of doppler cells	32–512
Max. velocity resolution (ms^{-1})	< 0.5
Primary mission	Ship detection
Revisit time for entire coverage area	Equal to CIT
Secondary missions	Aircraft detection Remote sensing
Number of simultaneous targets tracked	> 200

accurate predictions are required. The difficulties are compounded when mixed paths are involved; that is, when part of the propagation path lies over land, as happens where islands are present in the coverage area.

Focusing attention here on over-the-horizon detection, Figure 20.40 shows how the surface wave decays with range, parametric in frequency for the case in which both the radar antenna and the target are near the sea surface. These curves are for a smooth surface and use a 4/3 earth radius to approximate atmospheric refraction effects. The propagation code used here is due to Berry and Chrisman,[157] and it is quite flexible, permitting antenna and target altitudes, surface conductivity and permittivity, polarization, and frequency to be specified. The key points to be drawn from this example are (i) the apparent advantage to be gained by operating at low frequencies where propagation losses are minimized, though this benefit must be balanced against antenna size, the higher noise environment, and (often) reduced target RCS; and (ii) the rapidly accelerating fall-off in signal strength at ranges beyond a few hundred kilometers, where 10 dB of extra transmit power may buy only on the order of 10 kilometers of additional detection range.

A more widely used propagation code is GRWAVE,[158] which employs different mathematical representations for the field, depending on range and other parameters, so as to maximize computational efficiency. Some indication of the accuracy of this model

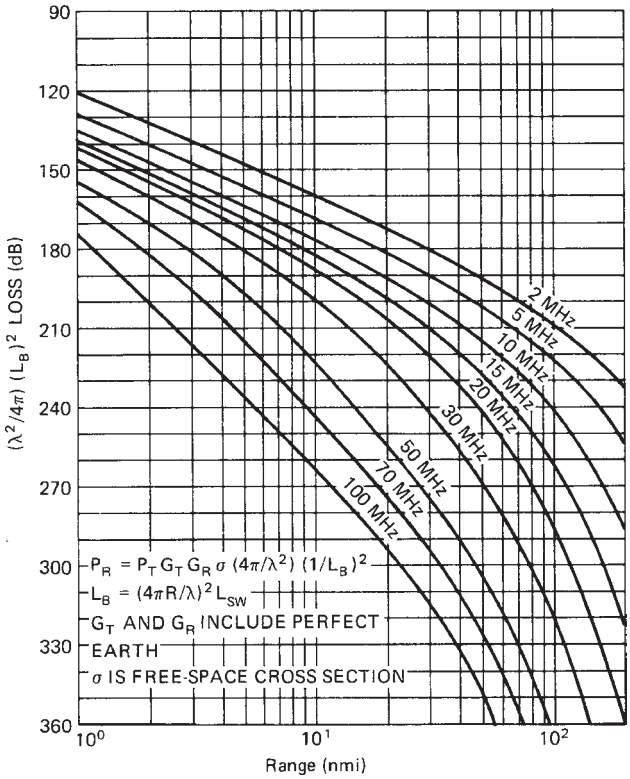


FIGURE 20.40 Curves of propagation loss versus range, as used for estimating surface wave radar performance. The surface is assumed to be smooth, target and antenna heights are 2 m, conductivity is taken as 5 S/m, and the dielectric constant is 80.

can be found in Figure 20.41, which compares GRWAVE predictions with experimental measurements of the signal strength received at a shore-based HFSWR when the transmitter was carried by a small boat traveling out to a range of ~110 km.[156] To avoid peripheral issues concerning antennas, the GRWAVE curves have been arbitrarily normalized here to align with the measurements at a range of 40 km. It is clear that, with this normalization, the predictions match reasonably well over the whole range extent, with small but systematic departures. GRWAVE seems to underestimate attenuation slightly at 7.72 MHz but over-estimate at 12.42 MHz. The sea roughness in this case was low (sea state 1–2). The impact of sea roughness on signal amplitude can be taken into account by using the expressions for roughness loss derived by Barrick.[159]

Another surface wave propagation modeling code has been developed by Sevgi,[160] paying particular attention to the calculation of propagation over hybrid paths with multiple islands.

The effect of the time-varying sea roughness on the signal phase and wavefront structure, manifested in the time delay, doppler spectrum, and direction-of-arrival spectrum of the received signal, can be computed using the multiple scattering theory of Anderson et al.[161]

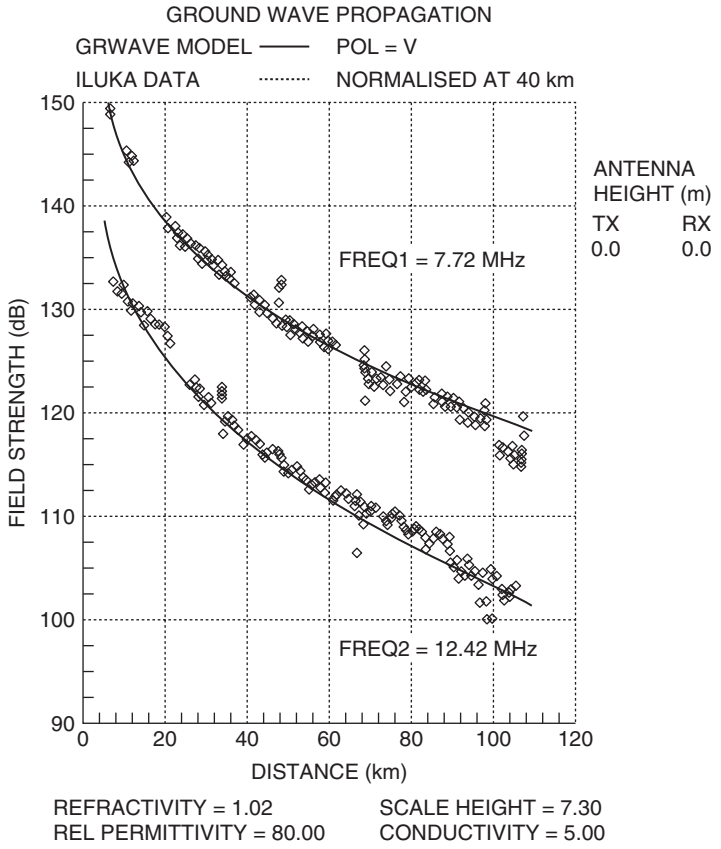


FIGURE 20.41 Experimental measurements of one-way surface wave attenuation, compared with GRWAVE predictions. Note the fluctuations due to scalloping loss as the source moved through FFT range bins.

Scattering: Targets and Clutter. The discussion of HF RCS and sea clutter in Sections 20.7 and 20.8 applies equally to HFSWR; indeed, the absence of the corrupting effects of the ionosphere provides even greater scope for exploiting the scattered signals. Furthermore, although the increased rate of decay of the surface wave at higher frequencies imposes a limit on the range of frequencies that might be used to detect a target at a given range, the situation is by no means as restrictive as for skywave propagation. This raises the possibility of exploiting multiple frequencies more effectively, so as to extract additional target and sea-state information and unmask targets hidden in clutter.

A compelling example is presented in Figure 20.42, where a range cell containing a ship target traveling at a speed of 13 knots has been interrogated at eight radar frequencies and the resulting doppler spectra plotted in a nested display.

The two columns of plots show received power versus doppler frequency for the eight operating frequencies as labeled, and for the target approaching (right column) and receding (left column). The abscissa units are in doppler normalized to

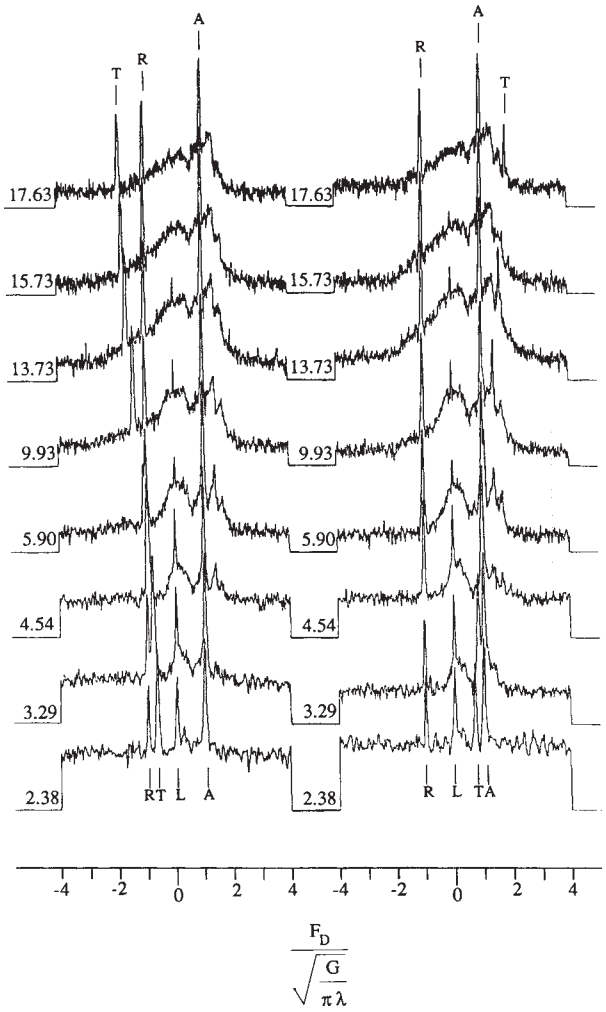


FIGURE 20.42 Multifrequency HFSWR doppler spectra showing the different frequency dependence of (i) target and (ii) clutter spectrum characteristics. The target T is shown when approaching (right) and receding (left) in the presence of the sea echo, with Bragg peaks marked A for approaching and R for receding. The curves plot received power versus normalized doppler (i.e., in units of the Bragg frequency) for eight radar operating frequencies. The peak at zero frequency is due to a stationary target in an antenna sidelobe.

the resonant wave or Bragg frequency; therefore, the resonant wave responses peak at ± 1. The amplitude range for each plot is 60 dB. The narrow peaks at zero doppler frequency are due to land in an antenna sidelobe. The target doppler and Bragg line frequency coincide at a radar frequency of 4.93 MHz; the target doppler lies between the Bragg lines for radar frequencies below that frequency and outside them for

frequencies above. The positive doppler resonant wave peak is about 20 dB larger than the negative doppler peak, indicating a sea driven by winds blowing toward the radar. The processing used in developing these displays was 200-s CIT and 30-min averaging.

Performance Modeling. Examples of HFSWR system performance modeling[160,162] generally follow the precepts described in Section 20.12. The main difference here is the availability of path-loss descriptors such as those shown in Figure 20.40. For example, consider a radar at 5 MHz with an average power of 10 kW (40 dBW), a transmit-receive antenna gain product of 15 dB, and a target at 100 nmi with an RCS of 20 dBsm; then the received power is

$$P_r = 40+15+20-222 = -147\,dBW$$

By using the January nighttime noise as given in Figure 20.19*b*,

$$SNR = P_r - N = -147+153 = 6\,dB$$

and if 10 s coherent processing time is used,

$$SNR = 16\,dB$$

As noted earlier, propagation loss accelerates rapidly with distance, especially at higher frequencies, while atmospheric and surface roughness effects will accumulate and contamination of the surface wave return with echoes received via skywave paths will become increasingly severe, so quantitative performance predictions at long ranges, beyond ~200 km, should be treated with caution.

REFERENCES

1. A. H. Taylor and E. O. Hulbert, "The propagation of radio waves over the earth," *Physical Review*, vol. 27, February 1926.
2. L. A. Gebhard, "Evolution of naval radio-electronics and contributions of the Naval Research Laboratory," Naval Res. Lab. Rept. 8300, 1979.
3. J. M. Headrick and M. I. Skolnik, "Over-the-Horizon Radar in the HF Band," *Proc. IEEE*, vol. 62, pp. 664–673, June 1974.
4. D. A. Boutacoff, "Backscatter radar extends early warning times," *Defense Electronics*., vol. 17, pp. 71–83, May 1985.
5. Guest editorial and invited papers in special issue on high-frequency and ice mapping and ship location, *IEEE J. Oceanic Eng*., vol. OE-11, April 1986.
6. J. R. Barnum, "Ship detection with high resolution HF skywave radar," *ibid.*, pp. 196–210, April 1986.
7. J. M. Headrick, "Looking over the horizon," *IEEE Spectrum*, vol. 27, pp. 36–39, July 1990.
8. D. H. Sinnott, "The Jindalee over-the-horizon radar system," *Conf. Air Power in the Defence of Australia*, Australian National University, Research School of Pacific Studies, Strategic and Defence Studies Centre, Canberra, Australia, July 14–18, 1986.
9. J. Wylder, "The frontier for sensor technology," *Signal*, vol. 41, pp. 73–76, 1987.
10. V. A. Yakunin, F. F. Evstratov, F. I. Shustov, V. A. Alebastrov, and Y. I. Abramovich, "Thirty years of eastern OTH radars: history, achievements and forecast," *L'Onde Electrique*, vol. 74, no. 3, May–June 1994.

11. C. Goutelard, "The NOSTRADAMUS project: French OTH-B radar design studies," *47th AGARD Symposium on 'Use or Reduction of Propagation and Noise Effects in Distributed Military Systems'*, AGARD CP-488 (Supp.), Greece, October 1990.
12. C. Goutelard, "STUDIO father of NOSTRADAMUS. Some considerations on the limits of detection possibilities of HF radars," *Int. Conf. HF Radio Systems and Techniques*, IEE Conference Publication no. 474, July 2000.
13. V. Bazin, J. P. Molinie, J. Munoz, P. Dorey, S. Saillant, G. Auffray, V. Rannou, and M. Lesturgie, "A general presentation about the OTH-Radar NOSTRADAMUS," *IEEE Radar Conference*, Syracuse, NY, May 2006. Also reprinted in *IEEE AES Systems Magazine*, vol. 21, no. 10, pp. 3–11, October 2006.
14. Zhou Wenyu and Mao Xu, "Bistatic FMCW OTH-B experimental radar," *Proc. Int. Conf. Radar ICR-91*, China Institute of Electronics, 1991, pp. 138–141.
15. Guest editorial and invited papers reviewing OTH radar technology, with emphasis on recent progress, *Radio Science*, vol. 33, July–August 1998.
16. A. A. Kolosov (ed.), *Fundamentals of Over-the-Horizon Radar*, in Russian, Radio i svyaz, 1984. Also a translation by W. F. Barton, Norwood, MA: Artech House, 1987.
17. R. A. Greenwald, K. B. Baker, R. A. Hutchins, and C. Hanuise, "An HF phased array radar for studying small-scale structure in the high latitude ionosphere," *Radio Science*, vol. 20, pp. 63–79, January–February 1985.
18. P. A. Bernhardt, G. Ganguli, M. C. Kelley, and W. E. Swartz, "Enhanced radar backscatter from space shuttle exhaust in the ionosphere," *J. Geophys. Res*., vol. 100, pp. 23,811–23,818, 1995.
19. R. M. Thomas, P. S. Whitham, and W. G. Elford, "Response of high frequency radar to meteor backscatter," *J. Atmos. Terr. Phys*., vol. 50, pp. 703–724, 1988.
20. A. Cameron, "The Jindalee operational radar network: its architecture and surveillance capability," *Proc. IEEE Int. Radar Conf*., 1995, pp. 692–697.
21. K. Davies, *Ionospheric Radio,* London: P. Peregrinus, 1990.
22. J. Thomason, G. Skaggs, and J. Lloyd, "A global ionospheric model," Naval Res. Lab. Rept. 8321, August 20, 1979.
23. K. Hocke and K. Schlegel, "A Review of atmospheric gravity waves and travelling ionospheric disturbances: 1982–1995," *Annales Geophysicae*, vol. 14, pp. 917–940, 1996.
24. A. Bourdillon, J. Delloue, and J. Parent, "Effects of geomagnetic pulsations on the doppler shift of HF backscatter radar echoes," *Radio Science*, vol. 24, pp. 183–195, 1989.
25. B.G. Fejer and M. C. Kelley, "Ionospheric irregularities," *Rev. Geophys. and Space Phys*., vol. 18, pp. 401–454, 1980.
26. C.-S. Huang, M. C. Kelley, and D. L. Hysell, "Nonlinear Rayleigh-Taylor instabilities, atmospheric gravity waves, and equatorial spread F," *J. Geophys. Res*., vol. 9, pp. 15,631–15,642, 1993.
27. D. L. Lucas and G. W. Haydon, "Predicting statistical performance indexes for high frequency telecommunications systems," ESSA Tech. Rept. IER 1 ITSA 1, U.S. Department of Commerce, 1966.
28. A. L. Barghausen, J. W. Finney, L. L. Proctor, and L. D. Schultz, "Predicting long-term operational parameters of high-frequency sky-wave communications systems," ESSA Tech. Rept. ERL 110-ITS 78, U.S. Department of Commerce, 1969.
29. J. M. Headrick, J. F. Thomason, D. L. Lucas, S. McCammon, R. Hanson, and J. L. Lloyd, "Virtual path tracing for HF Radar including an ionospheric model," Naval Res. Lab. Memo. Rept. 2226, March 1971.
30. L. R., Teters, J. L. Lloyd, G. W. Haydon, and D. L. Lucas, "Estimating the performance of telecommunication systems using the ionospheric transmission channel—ionospheric communications analysis and prediction program users manual," Nat. Telecom. Inf. Adm. NTIA Rept. 83–127, July 1983.
31. V. E. Hatfield, "HF communications predictions 1978 (An economical up-to-date computer code, AMBCOM)," *Solar Terrestrial Production Proc*., vol. 4, in *Prediction of Terrestrial Effects of Solar Activity*, R. F. Donnelley (ed.), National Oceanic and Atmospheric Administration, 1980.

32. D. Lucas, G. Pinson, and R. Pilon, "Some results of RADARC-2 equatorial spread doppler clutter predictions," *Proc. 7th Int. Ionospheric Effects Symp*., Alexandria, Virginia, pp. 2A5-1–2A5-8, May 1993.
33. D. L. Lucas, "Ionospheric parameters used in predicting the performance of high frequency skywave circuits," Interim Report on NRL Contract N00014-87-K-20009, Account 153-6943, University of Colorado, Boulder, April 15, 1987.
34. D. C. Miller and J. Gibbs, "Ionospheric analysis and ionospheric modeling," AFCRL Tech. Rept. 75-549, July 1975.
35. D. Bilitza, "International reference ionosphere," http://modelweb.gsfc.nasa.gov/ionos/iri.html.
36. Index of /models/ionospheric/iri/iri2001, http://nssdcftp.gsfc.nasa.gov/models/ionospheric/iri/iri2001.
37. A. G. Kim, Z. F. Zumbrava, V. P. Grozov, G. V. Kotovich, Y. S. Mikhaylov, and A. V. Oinats, "The correction technique for IRI model on the basis of oblique sounding data and simulation of ionospheric disturbance parameters," *Proc. XXVIIIth URSI General Assembly*, New Delhi, October 2005.
38. R. E. Daniell, Jr. and D. N. Anderson, "PIM model 1995," http://modelweb.gsfc.nasa.gov/ionos/pim.html.
39. Parameterized Ionospheric Model, Computational Physics, http://www.cpi.com/products/pim/.
40. R.E. Daniell, L.D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka, and R. W. Schunk, "Parameterized ionospheric model: A global ionospheric parameterization based on first principle models," *Radio Science*, vol. 30, pp. 1499–1510, 1995.
41. R. E. Daniell, "PRISM: assimilating disparate data types for improved low latitude ionospheric specification," presented at the Ionospheric Determination and Specification for Ocean Altimetry and GPS Surface Reflection Workshop at the Jet Propulsion Laboratory, Pasadena, CA, 2–4 December 1997.
42. D. N. Anderson, J. M. Forbes, and M. Codrescu, "A fully analytical, low- and middle-latitude ionospheric model," *J. Geophys. Res*. vol. 94, pp. 1520–1524, 1989.
43. Global Assimilation of Ionospheric Measurements, Park City, Utah, 2001, http://gaim.cass.usu.edu/GAIM/htdocs/present.htm.
44. Global Assimilative Ionospheric Model, JPL, http://iono.jpl.nasa.gov/gaim/index.html.
45. J. D. Huba, G. Joyce, and J. A. Fedder, "SAMI2 (Sami2 is another model of the ionosphere), A new low-latitude ionosphere model," *J. Geophys. Res*., vol. 105, 23,035–23053, 2000.
46. B. Khattatov, M. Murphy, M. Gnedin, T. Fuller- Rowell, and V. Yudin, "Advanced modeling of the ionosphere and upper atmosphere," *Environmental Research Technologies Report*, A550924, June 2004.
47. J. K. Hill, "Exact ray paths in a multisegment quasi-parabolic ionosphere," *Radio Science*, vol. 14, pp. 855–861, 1979.
48. T. A. Croft and H. Hoogasian, "Exact ray calculations in a quasi-parabolic ionosphere with no Magnetic Field," *Radio Science*, vol. 3, pp. 69–74, 1968.
49. R. J. Newton, P. L. Dyson, and J. A. Bennett, "Analytic ray parameters for the quasi-cubic segment model of the ionosphere," *Radio Science*, vol. 32, pp. 567–578, 1997.
50. R. M. Jones and J. J. Stephenson, "A versatile three-dimensional ray tracing computer program for radio waves in the ionosphere," Office Telecom. Rept. 75–76, October 1975.
51. C. J. Coleman, "A general purpose ionospheric ray-tracing procedure," DSTO Technical Report SRL-0131-TR, 1993.
52. Jari Perkiömäki, "High-frequency (HF) ionospheric communications propagation analysis and prediction," VOACAP Quick Guide, http://www.voacap.com/.
53. "Advanced stand alone prediction system," IPS Radio and Space Services, The Australian Space Weather Agency, http://www.ips.gov.au/Products_and_Services/1/1.
54. PROPLAB-PRO version 2.0, http://www.spacew.com/www/proplab.html.
55. B. T. Root and J. M. Headrick, "Comparison of RADARC High-frequency radar performance prediction model and ROTHR Amchitka data," Naval Res. Lab. Rept. NRL/MR/5320-93-7181, July 1993.

56. J. M. Headrick, B. T. Root, and J. F. Thomason, "RADARC model comparisons with Amchitka radar data," *Radio Science*, vol. 30, pp. 729–737, May–June 1995.
57. "New wind model," HWM 93, http://nssdcftp.gsfc.nasa.gov/models/atmospheric/hwm93/hwm93.txt.
58. J. A. Secan, R. M. Bussey, E. J. Fremouw, and Sa. Basu, "An improved model of equatorial scintillation," *Radio Science, 30,* 607–617, 1995.
59. A. V. Gurevich, *Nonlinear Phenomena in the Ionosphere,* New York: Springer-Verlag, 1978.
60. V. A. Alebastrov, A. T. Mal'tsev, V. M. Oros, A. G. Shlionskiy, and O. I. Yarko, "Some characteristics of echo signals," *Telecomm. and Radio Eng*., vol. 48, pp. 92–95, 1993.
61. V. G. Somov, V. A. Leusenko, V. N. Tyapkin, and G. Ya. Shaidurov, "Effect of nonlinear and focusing ionospheric properties on qualitative characteristics of radar in the decametric-wave band," *J. Comm. Technology and Electronics*, vol. 48, pp. 850–858, 2003.
62. ITT Avionics Division, Electro-Physics Laboratories, EPL Model ATL-75 Transmitter for Radar and Communication, IR&D Project Rept. 274, Results of performance measurements, January 1975.
63. D. J. Hoft and Fuat Agi, "Solid state transmitters for modern radar applications," *CIE Int. Radar Conf. Record, Beijing,* November 4–7, 1986, pp. 775–781.
64. F. A. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Trans. Microwave Theor. and Tech.*, vol. 50, pp. 814–826, March 2002.
65. D. J. Netherway and Carson, C. T., "Impedance and scattering matrices of a wideband HF phased Array," *J. Electron. Eng. Aust*., vol. 6, pp. 29–39, 1986.
66. Guest editorial and invited papers in special issue on shortwave broadcasting, *IEEE Trans. Broadcast.,* vol. 34, June 1988.
67. R. C. Johnson and H. Jasik (eds.), *Antenna Engineering Handbook,* 3rd Ed., New York: McGraw-Hill Book Company, 1993.
68. A. G. Kurashov (ed.), *Shortwave Antennas,* 2 Ed., in Russian, *Radio i svyaz,* January 1985.
69. S. J. Anderson, "Limits to the extraction of information from multi-hop skywave radar signals," *Proc. Int. Radar Conf*., Adelaide, September 2003, pp. 497–503.
70. S. J. Anderson, "The doppler structure of diffusely-scattered skywave radar echoes," *Proc. Int. Radar Conf*., Toulouse, October 2004.
71. S. J. Anderson, "Target classification, recognition and identification with HF radar, proc. NATO Research and Technology Agency," *Sensors and Electronics Technology Panel Symposium SET-080/RSY17/RFT 'TARGET IDENTIFICATION AND RECOGNITION USING RF SYSTEMS'*, Oslo, Norway, October 2004.
72. E. K. Walton and J. D. Young, "The Ohio State University compact radar cross section measurement range," *IEEE Trans. Ant. Prop.,* vol. AP-32, pp. 1218–1223, November 1984.
73. G. J. Burke and A. J. Poggio, "Numerical electromagnetic code (nec)-method of moments," NOSC Tech. Doc. 116, 1981.
74. R. W. Bogle and D. B. Trizna, "Small boat radar cross sections," Naval Res. Lab. Memo. Rept. 3322, July 1976.
75. R. Dinger, E. Nelson, S. Anderson, F. Earl, and M. Tyler, "High frequency radar cross section measurements of surrogate go-fast boats in Darwin, Australia," *SPAWAR System Center Tech. Rept. 1805*, September 1999.
76. S. J. Anderson, "Remote sensing with the Jindalee Skywave Radar," *IEEE J. Ocean. Eng.,* vol. OE- II, pp. 158–163, April 1986.
77. J. R. Barnum and E. E. Simpson, "Over-the-horizon radar target registration improvement by terrain feature localization," *Radio Science*, vol. 33, pp. 1067, July–August 1998.
78. D. E., Barrick, J. M. Headrick, R. W. Bogle, and D. D. Crombie, "Sea backscatter at HF: Interpretation and utilization of the echo," *Proc. IEEE,* vol. 62, pp. 673–680, June 1974.
79. S. O. Rice, "Reflection of electromagnetic waves from slightly rough surfaces," in *Theory of Electromagnetic Waves,* M. Kline (ed.), New York: Interscience Publishers, 1951, pp. 351–378.

80. D. E. Barrick, "First order theory and analysis of MF/HF/VHF scatter from the sea," *IEEE Trans.,* vol. AP-20, pp. 2–10, January 1972.
81. D. E. Barrick, "Remote sensing of sea state by radar," Chapter 12 in *Remote Sensing of the Troposphere*, V.E. Derr (ed.), Boulder, CO: NOAA/Environmental Research Laboratories, 1972, pp. 12.1–12.6.
82. D. D. Crombie, "Doppler spectrum of the sea echo at 13.56 Mcs," *Nature,* vol. 175, pp. 681–682, 1955.
83. J. W. Maresca, Jr. and J.R. Barnum, "Theoretical limitation of the sea on the detection of low doppler targets by over-the-horizon radar," *IEEE Trans. Ant. Prop*., vol. AP-30, pp. 837–845, 1982.
84. W. J. Pierson and L. Moskowitz, "A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigordskii," *J. Geophys. Res.,* vol. 69, no. 24, pp. 5181–5190, 1964.
85. K. Hasselmann, D. B. Ross, P. Muller, and W. Sell, "A parametric wave prediction model," *J. Phys. Oceanogr,* vol. 6, pp. 200–228, 1976.
86. T. Elfouhaily, B. Chapron, K. Katsaros, and D. Vandemark, "A unified directional spectrum for long and short wind-driven waves," *J. Geophys. Res.,* vol. 102, pp. 15782–15796, 1997.
87. A. E. Long and D. B. Trizna, "Mapping of North Atlantic winds by HF radar sea backscatter interpretation," *IEEE Trans. Ant. Prop.,* vol. AP-21, pp. 680–685, September 1973.
88. L. R. Wyatt, "A relaxation method for integral inversion applied to HF radar measurement of the ocean wave directional spectrum," *Int. J. Remote Sens.,* vol. 11, pp. 1481–1494, August 1990.
89. Y. Hisaki, "Nonlinear inversion of the integral equation to estimate ocean wave spectra from HF radar," *Radio Science,* vol. 31, pp. 25–39, 1996.
90. N. Hashimoto and M. Tokuda, "A Bayesian approach for estimation of directional wave spectra with HF radar," *Coastal Eng. J.*, vol. 41, pp.137–149, 1999.
91. D. E. Barrick, "Extraction of wave parameters from measured hf radar sea-echo spectra," *Radio Science,* vol, 12, no. 3, p. 415, 1977.
92. T. M. Georges, J. A. Harlan, R. R. Leben, and R. A. Lematta, "A test of ocean surface current mapping with over-the-horizon radar," *IEEE Trans. Geosci. and Rem. Sens.*, vol. 36. pp. 101–110, 1998.
93. H. L. Tolman, WAVEWATCH III, National Weather Service, http://polar.ncep.noaa.gov/waves/wavewatch/wavewatch.html.
94. J. L. Ahearn, S. R. Curley, J. M. Headrick, and D. B. Trizna, "Tests of remote skywave measurement of ocean surface conditions," *Proc. IEEE,* vol. 62, pp. 681–686, June 1974.
95. D. B. Trizna and J. M. Headrick, "Ionospheric effects on HF over-the-horizon radar," in Goodman, J. M. (ed.), *Proc. Effect Ionosphere on Radiowave Systems,* ONR/AFGL-sponsored, April 14–16, 1961, pp. 262–272.
96. J. Parent and A. Bourdillon, "A Method to correct HF skywave backscattered signals for ionospheric frequency modulation," *IEEE Trans. Ant. Prop.*, vol. AP-36, pp. 127–135, 1987.
97. S. J. Anderson and Y.I. Abramovich, "A unified approach to detection, classification and correction of ionospheric distortion in HF skywave radar systems," *Radio Science*, vol. 33, pp.1055–1067, July–August 1998.
98. D. B. Trizna, "Estimation of the sea surface radar cross section at HF from second-order doppler spectrum characteristics," Naval Res. Lab. Rept. 8579, May 1982.
99. R. O. Pilon and J. M. Headrick, "Estimating the scattering coefficient of the ocean surface for high-frequency over-the-horizon radar," Naval Res. Lab. Memo. Rept. 5741, May 1986.
100. J. Jones and P. Brown, "Sporadic meteor radiant distributions: orbital survey results," *Mon. Not. Roy. Astr. Soc*., vol. 265, pp. 524–-532, 1993.
101. M. A. Cervera and W. G. Elford., "The meteor response function: theory and application to narrow beam MST radar," *Planet. Space Sci.*, vol. 52, pp. 591–602, 2004.
102. P. Brown and J. Jones, "A determination of the strengths of the sporadic radio-meteor sources," *Earth, Moon and Planets*, vol. 68, pp. 223–-245, 1995.

103. M. A. Cervera, D. A. Holdsworth, I. M. Reid, and M. Tsutsumi, "The meteor radar response function: Application to the interpretation of meteor backscatter at medium frequency," *J. Geophys. Res*., vol. A109, pp. 11309, 2004.
104. T. J. Elkins, "A model for high frequency radar auroral clutter," RADC Rept. TR-80-122, March 1980.
105. "World distribution and characteristics of atmospheric radio noise," CCIR Rept. 322, CCIR (International Radio Consultative Committee), International Telecommunications Union, editions 1964, 1983, and 1988.
106. A. D. Spaulding and J. S. Washburn, "Atmospheric radio noise: Worldwide levels and other characteristics," NTIA Rept. 85-173, National Telecommunications and Information Administration, April 1985.
107. D. L. Lucas and J. D. Harper, "A numerical representation of CCIR Report 322 high frequency (3-30 Mcs) atmospheric radio noise data," Nat. Bur. Stand. Note 318, August 5, 1965.
108. D. B. Sailors, "Discrepancy in the International Radio Consultative Committee Report 322-3 radio noise model: The probable cause," *Radio Science*, vol. 30, pp. 713–728, 1995.
109. B. J. Northey and P. S. Whitham, "A comparison of DSTO and DERA HF background Noise measuring systems with the International Radio Consultative Committee (CCIR) model data," DSTO Technical Report DSTO-TR-0855, November 2000.
110. M. Kotaki and C. Katoh, "The global distribution of thunderstorm activity observed by the ionosphere satellite (ISS-b)," *J.Atmos. Terr. Phys*., vol. 45, pp. 833–850, 1984.
111. C. J. Coleman, "A direction-sensitive model of atmospheric noise and its application to the analysis of HF receiving antennas," *Radio Science*, vol. 37, pp. 3.1–3.10, 2002.
112. Yu. I. Abramovich, N. K. Spencer, and S. J. Anderson, "Experimental study of the spatial dynamics of environmental noise for a surface-wave OTHR application," *Proc. 8th Int. Conf. HF Radio Systems and Techniques*, IEE Conference Publication no. 474, Guildford, UK, July 2000, pp. 357–362.
113. L. E. Sweeney, "Spatial properties of ionospheric radio propagation as determined with half-degree azimuthal resolution," Stanford Electron. Lab. Tech. Rept. 155 SU- SEL-70-034, Stanford University, June 1970.
114. J. T. Lynch, "Aperture synthesis for HF radio signals propagated via the F-layer of the ionosphere," Stanford Electron. Lab. Tech. Rept. 161 SU-SEL-70-066, Stanford University, September 1970.
115. D. H. Sinnott and G. R. Haack, "The use of overlapped subarray techniques in simultaneous receive beam arrays," *Proc. Antenna Appl. Symp.*, University of Illinois, 1983.
116. S. J. Anderson, Y. I. Abramovich, and W-M. Boerner, "Measuring polarization dynamics of the generalized HF skywave channel transfer function," *Proc. Int Symp. Ant. and Prop.*, ISAP 2000, Japan, August 2000.
117. T. H. Pearce, "Receiving array design for over-the-horizon radar," *GEC J. Technology*, vol. 15, pp. 47–55, 1998.
118. G. F. Earl and M. J. Whitington, "HF radar ADC dynamic range requirements," *3rd Int. Conf. on Advanced A/D and D/A Conversion Techniques*, July 1999.
119. G. F. Earl, "FMCW waveform generator requirements for ionospheric over-the-horizon radar," *Radio Science*, vol. 33, pp. 1069–1076, 1998.
120. G. F. Earl, "Receiving system linearity requirements for HF radar," *IEEE Trans. Instrum. Meas*., vol. 40, pp. 1038–1041, 1991.
121. G. F. Earl, P. C. Kerr, and P. M. Roberts, "OTH radar receiving system design using synoptic HF environmental database," *Proc. 5th Int. Conf. HF Radio Systems and Techniques*, July 1991, pp. 48–53.
122. S. J. Anderson, "Simulation and modeling for the Jindalee over-the-horizon radar," *Math. and Comp. in Simulation*, vol. 27, pp. 241–248, 1985.
123. G. F. Earl, "Consideration of reciprocal mixing in HF OTH radar design," *Proc. 7th Int. Conf. HF Radio Systems and Techniques*, IEE Conference Publication no. 441, July 1997, pp. 256–259.

124. G. F. Earl, "HF radar receiving system image rejection requirements," *Proc. 6th Int. Conf. HF Radio Systems and Techniques*, September 1995, pp. 128–132.
125. T. H. Pearce, "Calibration of a large receiving array for HF radar," *Proc. Int. Conf. HF Radio Systems and Techniques*, IEE Conference Publication No. 411, July 1997, pp. 260–264.
126. D. M. Fernandez J. Vesecky, and C. Teague, "Calibration of HF radar systems with ships of opportunity," *Proc. of the 2003 IEEE Int. Geoscience and Remote Sensing Symp*., New York, July 2003, pp. 4271–4273.
127. I. S. D. Solomon, D. A. Gray, Y. I. Abramovich, and S. J. Anderson, "Over-the-horizon radar array calibration using echoes from ionized meteor trails," *IEE Proc. Radar, Sonar, and Navigation*, vol. 145, pp. 173–180, June 1998.
128. I. S. D. Solomon, D. A. Gray, Y. I. Abramovich, and S. J. Anderson, "Receiver array calibration using disparate sources," *IEEE Trans. Ant. Prop.*, vol. 47, pp. 496–505, March 1999.
129. G. J. Frazer and Y. I. Abramovich, "Quantifying multi-channel receiver calibration," DSTO Technical Report DSTO-TR-1152, 2001.
130. G. J. Frazer and S. J. Anderson, "Wigner-Ville analysis of HF radar measurements of an accelerating target," *Proc. 5th Int. Symp. Signal Proc. Appl.*, Brisbane, August 1999, pp. 317–320.
131. Y. Zhang, M. G. Amin, and G. J. Frazer, "High-resolution time-frequency distributions for maneuvering target detection in over-the-horizon radars," *IEE Proc. Radar, Sonar, and Navigation*, vol. 150, pp. 299–304, 2003.
132. T. Thayaparan and S. Kennedy, "Detection of a maneuvering air target in sea-clutter using joint time-frequency analysis techniques," *IEE Proc. Radar, Sonar, and Navigation*, vol. 151, pp. 19–30, February 2004.
133. G. J. Frazer and S. J. Anderson, "Estimating the frequency interval of a regularly spaced multicomponent harmonic line signal in colored noise," in *Defence Applications of Signal Processing*, D. A. Cochran, B. Moran, and L. White (eds.), New York: Elsevier, 2001, pp. 76–86.
134. G. Fabrizio, L. Scharf, A. Farina, and M. Turley, "Ship detection with HF surface-wave radar using short integration times," *Proc. Int. Conf. Radar 2004*, Toulouse, 2004.
135. D. O. Carhoun, J. D. R. Kramer Jr., and P. K. Rashogi, "Adaptive cancellation of atmospheric noise and ionospheric clutter for high frequency radar," MITRE Report, MTR 95B0000112, September 1995.
136. Y. I. Abramovich, S. J. Anderson, A. Y. Gorokhov, and N. K. Spencer, "Stochastically constrained spatial and spatio-temporal adaptive processing for nonstationary hot-clutter cancellation," in *Applications of Space-time Adaptive Processing*, R. K. Klemm (ed.), London: Springer, 2004, pp. 603–697.
137. Y. I. Abramovich, S. J. Anderson, A. Y. Gorokhov, and N. K. Spencer, "Stochastic constraints method in nonstationary hot clutter cancellation, part I: Fundamentals and supervised training applications," *IEEE Trans. AES*, vol. 34, pp. 1271–1292, October 1998.
138. Y. I. Abramovich, S. J. Anderson, and N. K. Spencer, "Stochastic-constraints method in nonstationary hot clutter cancellation, part II: Unsupervised training applications," *IEEE Trans. AES*, vol. 36, pp. 132–149, January 2000.
139. Y. I. Abramovich, V. N. Mikhaylyukov, and I. P. Malyavin, "Stabilisation of the autoregressive characteristics of spatial clutters in the case of nonstationary spatial filtering," *Sov. J. Commun. Technol. Electron*., vol. 37, pp. 10–19, 1992, translation of *Radioteknika I Electronika*.
140. R. Anderson, S. Kraut, and J. L. Krolik, "Robust altitude estimation for over-the-horizon radar using a state-space multipath fading model," *IEEE Trans. AES*, vol. 39, pp. 192–201, January 2003.
141. D. J. Percival and K. A. B. White, "Multipath coordinate registration and track fusion for over-the-horizon radar," in *Defence Applications of Signal Processing,* D. A. Cochran, B. Moran, and L. White (eds.), Amsterdam: Elsevier, 2001, pp. 149–155.
142. Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, New York: Academic Press, January 1988.
143. S. B. Colegrove and S. J. Davey, "PDAF with multiple clutter regions and target models," *IEEE Trans. AES*, vol. 39, pp. 110–124, January 2003.

144. J. L. Krolik and R. H. Anderson, "Maximum likelihood coordinate registration for over-the-horizon radar," *IEEE Trans. Sig. Proc*., vol. 45, pp. 945–959, 1997.

145. M. G. Rutten and D. J. Percival, "Joint ionospheric and track target state estimation for multipath othr track fusion," *Proc. SPIE Conf. on Signal and Data Processing of Small Targets*, 2001, pp. 118–129.

146. R. H. Anderson and J. L. Krolik, "Track association for over-the-horizon radar with a statistical ionospheric model," *IEEE Trans. Sig. Proc*., vol. 50, pp. 2632–2643, November 2002.

147. G. W. Pulford, "OTHR multipath tracking with uncertain coordinate registration," *IEEE Trans. AES*, vol. 40, pp. 38–56, 2004.

148. S. J. Anderson, F. J. Mei, and J. Peinan, "Enhanced OTHR ship detection via dual frequency operation," *Proc. China Institute of Electronics Int. Conf. on Radar*, Beijing, October 2001.

149. G. F. Earl and B. D. Ward, "Frequency management support for remote sea-state sensing using the JINDALEE skywave radar," *IEEE J. of Oceanic Engr*., vol. OE-11, pp. 164–173, April 1986.

150. G. F. Earl and B. D. Ward, "The frequency management system of the Jindalee over-the-horizon backscatter HF radar," *Radio Science*, vol. 22, pp. 275–291, 1987.

151. R. Barnes, "Automated propagation advice for OTHR ship detection," *IEE Proc. Radar, Sonar, and Navigation*, vol. 143, pp. 53–63, February 1996.

152. D. L. Lucas, J. L. Lloyd, J. M. Headrick, and J. F. Thomason, "Computer techniques for planning and management of OTH radars," Naval Res. Lab. Memo. Rept. 2500, September 1972.

153. J. M. Headrick, "HF over-the-horizon radar," Chapter 24 in *Radar Handbook*, M. I. Skolnik (ed.), 2nd Ed., New York: McGraw-Hill, 1990.

154. J. M. Hudnall and S. W. Der, "HF-OTH radar performance results," Naval Res. Lab. Tech. Rept. NRL/MR/5325-93-7326, 1993.

155. R. Fante and S. Dhar, "A model for target detection with over-the-horizon radar," *IEEE Trans. AES*, vol. 26, pp. 68–83, January 1990.

156. L. A. Berry and M. E. Chrisman, "A FORTRAN program for calculation of ground wave propagation over homogeneous spherical earth for dipole antennas," Nat. Bur. Stand. Rept. 9178, 1966.

157. S. Rotheram, "Ground wave propagation, parts 1 and 2," *IEE Proc., Pt. F*, vol. 128, pp. 275–295, 1981.

158. S. J. Anderson, P. J. Edwards, P. Marrone, and Y. I. Abramovich, "Investigations with SECAR—A bistatic HF surface wave radar," *Proc. IEEE Int. Conf. on Radar, RADAR 2003*, Adelaide, September 2003.

159. D. E. Barrick, "Theory of HF and VHF propagation across the rough sea, pts. I and 2," *Radio Science,* vol. 6, pp. 517–533, May 1971.

160. L. Sevgi, *Complex Electromagnetic Problems and Numerical Simulation Approaches,* Hoboken, NJ: IEEE Press, 2003.

161. S. J. Anderson, J. Praschifka, and I. M. Fuks, "Multiple scattering of HF radiowaves propagating across the sea surface," *Waves in Random Media*, vol. 8, pp. 283–302, April 1998.

162. G. H. Millman and G. Nelson, "Surface wave HF radar for over-the-horizon detection," *Proc. IEEE Int. Radar Conf*., 1980, pp. 106–112.

Chapter 21
Ground Penetrating Radar

David Daniels

ERA Technology

21.1 INTRODUCTION

The terms *ground penetrating radar (GPR), ground probing radar, subsurface radar,* or *surface penetrating radar (SPR)* refer to a radar-based electromagnetic technique designed primarily for the location of objects or interfaces buried beneath the Earth's surface or located within a visually opaque structure. GPR is a successful example of the exploitation of ultrawideband radar and typically a GPR with a range of 1 m would operate over the range 0.3 GHz to 3.3 GHz.

Although GPR has many similarities to radar systems, there are some key differences, which need to be appreciated when comparing them with conventional radar systems. GPR systems are a special class of ultrawideband (UWB) radar system and can radiate energy in the range of frequencies from a few MHz up to 10 GHz with a bandwidth of up to a decade, but more usually 2–3 octaves. The typical average radiated power, integrated over the band of interest, may be in the order of a milliwatt, but the power per Hz may be as low as picowatts.

GPR is usually operated so that the target, which is within a lossy dielectric, is only a few wavelengths from the aperture of antenna. The total path losses within a few wavelengths may reach 100 dB or more depending on the material. Many GPR systems operate in a region where the wavelengths radiated are greater or in the same order of magnitude as the target dimensions. Thus, GPR operates between the Rayleigh region and Mie or resonance region of the target dimensions. This is very different from conventional radar systems where the target dimensions are much larger than the wavelength of the incident radiation, i.e., the optical region.

The technology of GPR is largely applications-oriented and the overall design philosophy, as well as the hardware, is usually dependent on the target type and the material of the target and its surroundings. GPR is vulnerable to extremely high levels of clutter at short ranges and this, rather than signal/noise recovery, is its major technical challenge. The system to be specified should take this into account. All these aspects pose special design problems for GPR, which is described in detail by Daniels.[1] This chapter is a summary of that material and is referenced by courtesy of the IEE.

A typical GPR system is shown in Figure 21.1 and consists of a pair of antennas, one for transmit and one for receive, connected to the transmitter receiver and processor and contained within a sealed enclosure, a battery and control processor, and display unit. The wheels drive a shaft encoder that triggers data acquisition and hence the display that is synchronised to the movement of the system. An example of the display,

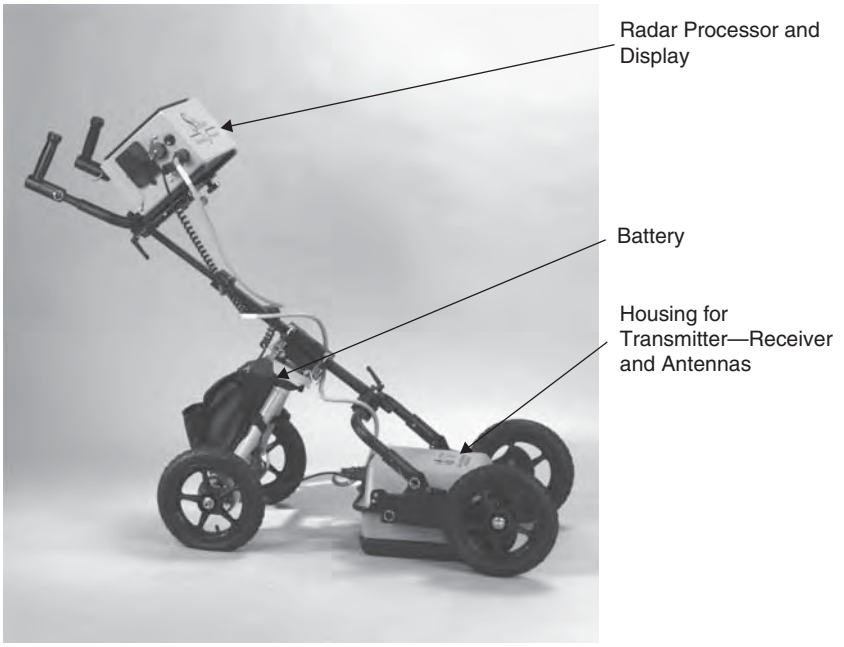


FIGURE 21.1 Typical GPR system (*Courtesy US Radar*)

which takes the form of a cross section of the ground surveyed by the GPR, is shown in Figure 21.2. The horizontal scale is 10 cm per marker and the vertical scale is time in nanoseconds (51 ns). An explanation of the image is provided later in this chapter.

GPR system design can be classified into two groups. GPR systems that transmit an impulse and receive the reflected signal from the target using a sampling receiver can be considered to operate in the time domain. GPR systems that transmit individual frequencies in a sequential manner and receive the reflected signal from the target using a frequency conversion receiver can be considered to operate in the frequency domain.

The first use of electromagnetic signals to determine the presence of remote terrestrial metal objects is generally attributed to Hülsmeyer in 1904, but the first description of their use for location of buried objects appeared six years later in a German patent by Leimbach and Löwy. The work of Hülsenbeck in 1926 appears to be the first use of pulsed techniques to determine the structure of buried features. He noted that any

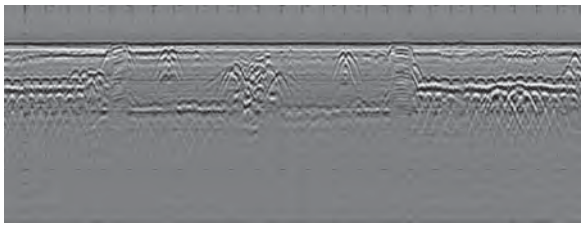


FIGURE 21.2 Typical display from GPR

dielectric variation, not necessarily involving conductivity, would also produce reflections and that the technique, through the easier realization of directional sources, had advantages over seismic methods. Pulsed techniques were developed from the 1930s onward as a means of probing to considerable depths in ice (Steenson[2] and Evans[3]); in fresh water and salt deposits (Unterberger[4]); in desert sand and rock formations (Kadaba[5] and Morey[6]). Probing of rock and coal was also investigated by Cook[7,8] as well as Roe,[9] although the higher attenuation in the latter material meant that depths greater than a few meters were impractical. Nilsson[10] gives a more extended account of the history of GPR and its growth up to the mid-1970s. From the 1970s, the range of applications has been expanding steadily, and now includes those given in Table 21.1. Purpose-built equipment for each of these applications has been being developed, and the user now has a better choice of equipment and techniques.

GPR has advanced rapidly as a result of a variety of applications, but as the requirements have become more demanding, so the equipment, techniques, and data processing methods have been developed and refined.

A GPR transmits a regular sequence of low-power pulses of electromagnetic energy into the material or ground and then receives and detects the weak reflected signal from the buried target. The energy is in the form of either a very short duration impulse, a sweep over a range of frequencies, radiation of noise over a defined band, or a pseudorandom coded sequence of pulses. Most GPR systems, all of which need to comply with the relevant national and international regulations regarding radio transmitters, operate within the range of frequencies from 10 MHz to 10 GHz and can have a bandwidth of several GHz. The FCC requirement for UWB limits the radiated power to –41 dBm Hz^{-1}. The topic of radar system design is covered in many texts and useful information will be found in the following references: Daniels[11,12] Cook and Bernfeld,[13] Skolnik,[14] Nathanson,[15] Wehner,[16] Galati,[17] and Astanin and Kostylev.[18]

The buried target can be a conductor, a dielectric, or combinations of both. The surrounding host material can be soil, earth materials, wood, rocks, ice, fresh water, or manmade materials such as concrete or brick. A typical GPR achieves a range of up to a few meters, but some special systems can penetrate up to hundreds of meters or even kilometers. A few GPR systems have been operated from aircraft and from satellites to image geological features buried beneath the Saharan deserts as well as measuring the depth of the Moon and features on Mars or comets. The range of the GPR in the ground is limited because of the absorption the signal undergoes, while it travels on its two-way path through the ground material. GPR works well through materials such as granite, dry sand, snow, ice, and fresh water, but will not penetrate certain clays that

TABLE 21.1 Main Applications of GPR

Archaeological investigations
Bridge deck analysis
Detection of buried mines (anti-personnel and anti-tank)
Forensic investigations (detecting buried bodies)
Geophysical investigations
Pipes and cable detection
Rail track and bed inspection
Road condition survey
Snow, ice, and glacier

are high in salt content or salt water because of the high absorption of electromagnetic energy of such materials. For comparison, the total two-way path loss from Earth to the Moon using a 1 GHz radar would be greater than 200 dB for a range of 356,400 km and a target radar cross section in the order of 10^{12} m^2, whereas a GPR radar often encounters a path loss in excess of 70 dB for ranges of less than a meter.

In air, the GPR signal travels at the speed of light, but is slowed down in ground materials by their dielectric constant; hence, true range needs calibrating for each material. GPR will not penetrate metal because of the latter's conductivity.

There are now a number of commercially available equipments, and the technique is gradually developing in scope and capability. Many GPR systems are mobile and mounted on wheels or skids to be moved by hand, but systems can be used on vehicles for rapid survey by means of an array of antennas. Other GPR systems are designed to be inserted into boreholes to provide images of the intervening rock. Typical GPR system attributes are given in Table 21.2.

Most GPR systems use separate, man-portable, transmit and receive antennas, which are placed on the surface of the ground and moved in a known pattern over the surface of the ground or material under investigation, and an image can be generated, in real time, on a display either in grey scale or in color. By systematically surveying the area in a regular grid pattern, a radar image of the ground can be built up. GPR images are displayed either as two-dimensional representations, using horizontal (x or y) and depth (z) axes or a horizontal plane representation (x, y) at a given depth (z) or as a three-dimensional reconstruction. GPR data may be classified as A-scan, B-scan, or C-scan depending on the plane of image (note these are not the same as conventional radar A, B, and C scans). A GPR A-scan is a measurement at a single fixed point in space and is displayed in amplitude (y) and range (x). A B-scan is a representation usually in grayscale or color-coded image intensity of a plane (x, z or y, z) of scan whereas a C-scan represents a horizontal plane (x, y) at a given depth (z). Alternatively, the GPR may be designed to provide an audible warning of target presence while the GPR is moved.

The GPR image of a target is very different from its optical image because the wavelengths of the illuminating radiation are similar in dimension to the target. This results in a much lower definition in the GPR image and one that is highly dependent on the propagation characteristics of the ground. The beam pattern of the antenna is widely spread in the dielectric and this degrades the spatial resolution of the image, unless corrected. Refraction and anisotropic characteristics of the ground may also distort the image. For some longer-range systems, synthetic aperture processing techniques are used to optimize the resolution of the image and will be discussed later.

TABLE 21.2 Characteristics of GPR Systems in a Soil of Relative Dielectric Constant of 9 and Loss Tangent of 0.1

Pulse Duration in ns	Center Frequency in MHz	Range in Meters	Depth Resolution
0.5	2000	<0.25	0.025
1	1000	<0.5	0.05
2	500	<1	0.1
4	250	<2	0.2
8	125	<4	0.4
16	63	<8	0.8
32	31	<16	1.6

Unprocessed GPR images often show "bright spots" caused by multiple internal reflections as well as a distortion of the aspect ratio of the image of the target caused by variations in the velocity of propagation. Symmetrical targets, such as spheres or pipes, cause migration of the reflected energy to a hyperbolic pattern. GPR images can be processed to compensate for these effects, and this is usually carried out offline. A GPR can be designed to detect specific targets such as interfaces in roads, pipes, and cables by means of polarized radiation and localized objects such as cubes, spheres, and cylinders. GPR is capable of detecting features many hundreds of years old; hence, a prospective site should remain unexcavated, prior to survey, so as to preserve its information.

A simplified diagram of the various sources of clutter in a GPR environment is given in Figure 21.3, and it can be seen that separation of the various signals is the key to identifying the wanted signal.

Inevitably there have been some claims for GPR capability that are simply outside the realms of known physics, and these seem to have been seized on by some sections of the media. A claim was made that a particular GPR and its operator could detect targets the size of golf balls at a depth of eight meters. Clearly, the wavelengths capable of propagating to eight meters in soil would be so much larger than a golf ball–sized target that the radar cross-sectional area of the latter would fade into insignificance, even noise. The persuasiveness of the claimant and the lack of understanding of basic physics on the part of some potential users enabled this kind of claim to be seriously considered. Claims were made that a GPR had been developed "that can provide three-dimensional images of objects up to 45.7 meters below the surface of land and sea. Such a device would allow verifiers to identify underground weapons facilities, like those of concern in Libya, Iraq, and North Korea. The underwater detection capability could also be used to verify treaties dealing with submarines and nuclear weapons positioned on the seabed." How well

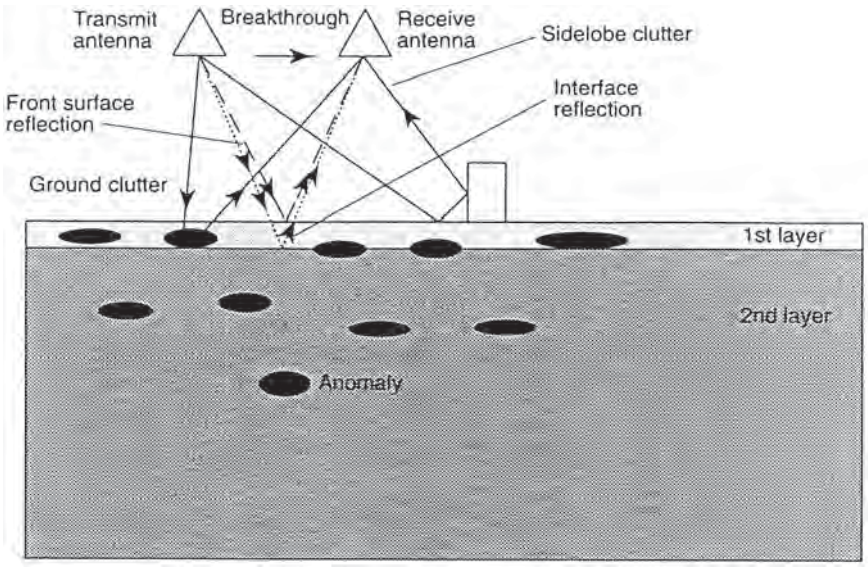


FIGURE 21.3 General system operation of GPR showing targets and sources of clutter (*Courtesy IEE*)

GPR would propagate through seawater is an interesting question given the known attenuation of seawater at radar frequencies. A careful analysis of some of the claims about the same radar was published by Tuley[19] and is interesting reading.

21.2 PHYSICS OF PROPAGATION IN MATERIALS

Introduction. Conventional radar systems are generally not significantly affected by the propagation characteristics of the medium the radar signals travel through, apart from rain, absorption spectra of the atmosphere, or ionized atmospheric layers. This is definitely not the case with GPR where the transmission medium may be non-isotropic, high dielectric and high loss, and may be layered. Therefore, an understanding of soil and material propagation characteristics is important, and this section describes the key features of the physics of propagation in materials.

Maxwell's equations are the foundation for the consideration of the propagation of electromagnetic waves. In free space, the magnetic susceptibility and electric permittivity are constants, that is, they are independent of frequency and the medium is not dispersive. In a dielectric with a zero loss tangent, no losses due to attenuation are encountered, and hence there is no consideration of the attenuation, which occurs in real dielectric media.

If an alternating electric field is applied to a material, the individual molecules will be induced to rotate in an oscillatory manner about an axis through their centers, the inertia of the molecules preventing them from responding instantaneously. Similar translational effects can occur. The polarization produced by an applied field (such as a propagating radar wave) is closely related to the thermal mobility of the molecules and is, therefore, strongly temperature dependent. Note that polarization in this context is different from the polarization of EM waves. In general, the relaxation time (which may be expressed as a relaxation frequency) depends on activation energy, the natural frequency of oscillation of the polarized particles, and on temperature. Relaxation frequencies vary widely between different materials.

For example, maximum absorption occurs at very low frequencies in ice (10^3 Hz), whereas it takes place in the microwave region in water (10^6 Hz–10^{10} Hz); thus, the effects of this phenomenon can have a direct bearing upon the dielectric properties of materials at the frequencies employed by GPRs, especially if moisture is present within a material. There are a number of other mechanisms, which cause a separation of positively and negatively charged ions resulting in electric polarization. These mechanisms can be associated with ionic atmospheres surrounding colloidal particles (particularly clay minerals), absorbed water, and pore effects, as well as interfacial phenomenon between particles.

The general form of the model that describes the frequency dependence of such systems is the Debye[19] relaxation equation:

$$\varepsilon' - i\varepsilon'' = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + i\omega\tau} \tag{21.1}$$

where

ε' is the real part of the dielectric permittivity,
ε' = real part of the dielectric permittivity
ε'' = imaginary part of the dielectric permittivity
ε_∞ = high frequency limiting value of the permittivity

ε_s = low frequency limiting value of the permittivity
ω = radian frequency = $2\pi f$)
τ = relaxation time constant

The frequency of maximum movement and loss occurs at $\omega = 1/\tau$.

In general, single relaxations are rarely observed in natural systems. Instead, there are distributions of relaxations corresponding to distributions of size scales that influence movement of charge. There are several equations describing such distributed systems, with the most common experimental observations in agreement with the model from Cole and Cole[20]:

$$\varepsilon' - i\varepsilon'' = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + (i\omega\tau)^\alpha} \tag{21.2}$$

where α describes the breadth of the time constant distribution from a single relaxation, $\alpha = 1$, to an infinitely broad distribution, $\alpha = 0$, with a common process. Different polarization processes may be described by a series of Cole-Cole equations with different values of α and other parameters.

The electromagnetic properties of a buried target must be different from the surrounding soil or material, and this means that to a first order its relative dielectric constant should be significantly lesser or greater than the host soil. Typically, most soils exhibit a relative dielectric constant, which ranges between 2 to 25. Fresh water has a relative dielectric constant of approximately 80. It should be noted that the ground and surface are quite likely to be inhomogeneous and contain inclusions of other rocks of various size as well as manmade debris. This suggests that the signal to clutter performance of the sensor is likely to be an important performance factor. Clutter may be regarded as any radar return that is not associated with the wanted target and needs to be defined with respect to a particular application.

Attenuation. Electromagnetic waves propagating through natural media experience losses, to both the electric (E) or magnetic (H) fields. This causes attenuation of the original electromagnetic wave. Plane waves are good approximations to real waves in many practical situations. More complicated electromagnetic wavefronts can be considered as a superimposition of plane waves, and this method may be used to gain an insight into more complex situations. For most soils of interest in GPR, the magnetic response is weak and need not be considered as a complex quantity, unlike the permittivity and conductivity. However, in certain soil types, such as those derived from volcanic rocks or otherwise high in iron content, full consideration of the magnetic properties is necessary. In the case of lossy dielectric materials, both conduction and dielectric effects cause absorption of electromagnetic radiation.

The electromagnetic material properties that describe such a system are in the complex propagation constant γ:

$$\gamma = ik = \alpha + i\beta \tag{21.3}$$

where γ = propagation constant
k = wave number ($2\pi/\lambda$)
α = attenuation constant [nepers/m]
β = phase constant [radians/m]

The field at a distance z from the source is given by

$$E(z,t) = E_0 \cdot e^{-\alpha \cdot z} \cdot e^{j(\omega t - \beta \cdot z)} \tag{21.4}$$

The wavelength, λ, in the medium is in meters,

$$\lambda = \frac{2\pi}{\beta} = \frac{v}{f} \tag{21.5}$$

where f is frequency in Hertz.

The losses in such systems are described in terms of tangents of loss angles, δ, between electric and magnetic fields. The electrical loss tangent is given by

$$\tan\delta_e = \frac{\varepsilon''}{\varepsilon'} + \frac{\sigma}{\omega\varepsilon'}, \text{ which can be simplified to } \tan\delta_e \approx \frac{\sigma}{\omega\varepsilon'} \text{ for low loss materials} \tag{21.6}$$

representing the sum of the charge transport and polarization relaxation losses, and the phase angle between electric field and current density. The skin depth or attenuation length is $1/\alpha$ [m]; the distance electromagnetic energy travels while being attenuated by 1/e in amplitude. This distance is known as the skin depth, d, and provides an initial guide to the useful penetration depth of a GPR system although in some media the useful range may be greater.

The individual propagation constants can be written as

$$\alpha = \omega\sqrt{\left[\frac{\mu\varepsilon'}{2}\sqrt{1+\left(\frac{\varepsilon''}{\varepsilon'}\right)^2} - 1\right]} \quad \beta = \omega\sqrt{\left[\frac{\mu\varepsilon'}{2}\sqrt{1+\left(\frac{\varepsilon''}{\varepsilon'}\right)^2} + 1\right]}$$

where

α = attenuation factor
β = phase constant

and the dimensionless factor $\varepsilon''/\varepsilon'$ is more commonly termed the material loss tangent.

This discussion has not considered the electromagnetic and magnetic loss tangent, and these may need to be considered in special cases.

It can be seen from the above expressions that the attenuation constant of a material is, to a first order, linearly related (in dBm^{-1}) to frequency. It is not sufficient to consider only the low frequency conductivity when attempting to determine the loss tangent over the frequency range 10^7 to 10^{10} Hz. In the case of a material that is dry and relatively lossless, it may be reasonable to consider that tan δ_e is constant over that frequency range. However, for materials that are wet and lossy such an approximation is invalid. However, there are a number of other factors that influence the effective penetration depth, notably the strength of reflection from the target sought, and the degree of clutter suppression of which the system is capable.

A first order estimate of the various contributions to signal loss can be carried out using the standard radar range equation, although this is only applicable for far-field conditions and thus has restrictions.

$$P_r = \frac{P_t.A.G.\sigma.k}{(4.\pi.R^2)^2}.e^{-\alpha 2R} \tag{21.7}$$

where

P_t = transmitted power in watts
P_r = received power in watts
A = antenna gain
G = antenna effective aperture
R = range in meters
α = target radar cross section
k = calibration coefficient

The cumulative losses include the transmission coefficients into the ground; the spreading losses describe the R^{-4} losses for a target of 1 m^2; and the attenuation losses are for a soil with a ε_r of 9 and tan δ of 0.1. Fixed losses include the transmission losses into the soil and the effective radar cross section of the target, which comprises its true radar cross section and reflection loss from the target. Note that a conducting reflector will have low return loss whereas a nonconducting reflector will have a high return loss. In Figure 21.4, the calculation has been derived from 1 meter to 10 meters as the radar range equation is not an accurate model in this range less than 1 meter. and the purpose of the explanation is to provide a basic introduction to first order signal estimation.

Reflection. In any estimation of received signal level, it is necessary to consider the coefficients of reflection and transmission, as the wave passes through the dielectric to the target and Snell's Laws describe the associated angles of incidence, reflection, transmission, and refraction. Where lossy materials are involved, complex angles of refraction may occur unlike the simple classical case, and polarization and the Stoke's matrix may also be required for oriented high-aspect ratio features like pipes, wires, and fractures.

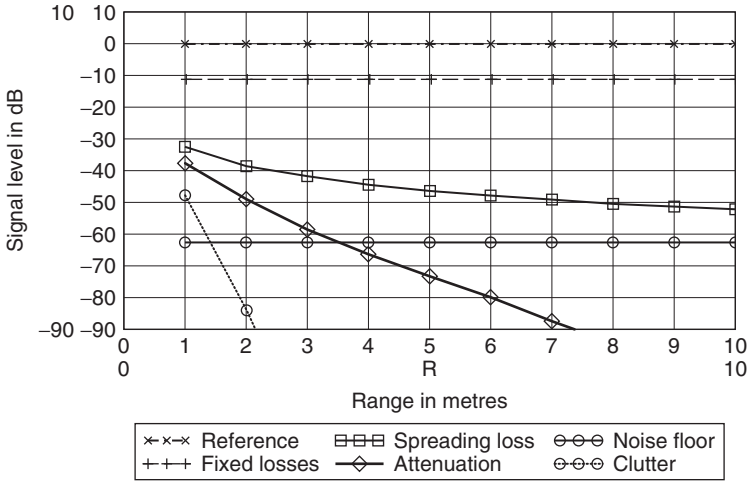


FIGURE 21.4 Losses for GPR signal versus range *(Courtesy IEE)*

The intrinsic impedance η of a medium is the relationship between the electric field, E, and the magnetic field, H, and is a complex quantity given by

$$\eta = \sqrt{\frac{-j\omega\mu}{\sigma - j\omega\varepsilon}} \tag{21.8}$$

At the boundary between two media, some energy will be reflected and the remainder transmitted. The reflected field strength is described by the reflection coefficient, r:

$$r = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} \tag{21.9}$$

where η_1 and η_2 are the impedances of medium 1 and 2, respectively.

The reflection coefficient has a positive value when $\eta_2 > \eta_1$, such as where an air-filled void exists in a dielectric material. The effect on a pulse waveform is to change the phase of the reflected wavelet so that targets with different relative dielectric constants to the host material show different phase patterns of the reflected signal. However, the propagation parameters (relative dielectric constant and loss tangent) of the host material, the geometric characteristics of the target, and its dielectric parameters affect the amplitude of the reflected signal.

Clutter. A major difficulty for operation of GPR systems is the presence of clutter within the material. *Clutter* is defined as sources of unwanted reflections that occur within the effective bandwidth and search window of the radar and present as spatially coherent reflectors. The definition of clutter very much depends on the wanted target. The operator of a GPR system searching for pipes may classify the interfaces between road layers as clutter, whereas the operator of a system measuring road layer thickness might consider pipes and cables as sources of clutter. Careful definition and understanding are critically important in selecting and operating the best system and processing algorithms. Clutter can completely obscure the buried target and a proper understanding of its source and impact on the radar is essential.

Polarization. A complete description of the radar scattering cross section of a target includes a description of its polarization scattering characteristics (not the same as molecular polarization). The polarizing properties of targets are described by the Stokes parameters, and the polarization coordinates can be represented on the Poincare Sphere. All of these are well described in standard texts on optics and electromagnetic theory. In summary, these descriptions allow the state of an electromagnetic wave to be described in terms of linear, elliptical, and circular polarization (left-handed or right-handed). It is well known that linear targets such as wires act as depolarizing features and that a linearly polarized crossed dipole antenna rotated about an axis normal to a linear target such as a wire or pipe produces a sinusoidal variation in received signal. However, the null points are a distinct disadvantage, because the operator is required to make two separate, axially rotated measurements at every point to be sure of detecting pipes at unknown orientations. An attractive technique is to radiate a circularly polarized wave, which automatically rotates the polarized vector in space and hence removes the direction of signal nulls. These techniques can be used to discriminate in favor of the target. For example, a right-handed circularly polarized (RHCP) wave will be reflected as a left-handed circularly

polarized (LHCP) wave from a planar surface, but some proportion of RHCP will be reflected from a thin pipe or wire. This enables the ground-surface reflection to be reduced while enhancing that from the thin pipe or wire.

Velocity. The velocity of propagation of electromagnetic waves in free space is approximately 3×10^8 ms^{-1} but slows in a material depending on its relative permittivity and relative magnetic permeability. The velocity of propagation of electromagnetic waves in a soil with a value for ε_r of 9 would be slowed to 1×10^8 ms^{-1}. The time to a target at a range of 1 meter is, therefore, 20 ns, and GPR systems operate at time ranges between a few nanoseconds up to 200 ns, although some systems for probing through ice may use ranges up to several tens of milliseconds.

In general, it is not possible to make a reliable estimate of propagation velocity or relative permittivity in a medium from a single measurement without trial holing (inserting a probe into a predrilled hole) or other supplementary information. Even in the case where a measurement is carried out at one location, it is often found that significant variations in velocity will occur within comparatively short distances from the original location. This can lead to significant errors in the estimation of depths of reflectors. One procedure that overcomes this limitation is known as common depth point surveying, which utilizes two antennas in bistatic operation at a number of transmit and receive positions.

The velocity of propagation is given by $v = (\mu_o \mu_r \varepsilon_o \varepsilon_r)^{-1/2}$; hence in a material with $\mu_r = 1$, the velocity becomes $v = c/(e_r)^{1/2}$.

The phase velocity is given by $v = \omega/\beta$ and as

$$\beta = \omega \sqrt{\left[\frac{\mu\varepsilon'}{2} \left(\sqrt{1 + \left(\frac{\varepsilon''}{\varepsilon'} \right)^2} + 1 \right) \right]} \tag{21.10}$$

The phase velocity is also dependent on the factor $\varepsilon''/\varepsilon'$, which is also tan δ.

It is also possible to derive velocity from multiple measurements scanning over a target, but this works well only in relatively uncluttered situations where the media has no anisotropic characteristics.

Dispersion. The frequency dependent nature of the dielectric properties of the material causes the phase velocity of the component frequencies of a wideband signal to suffer differential propagation values. Hence, there will be variation in the velocity of propagation with frequency. Dielectrics exhibiting this phenomenon are termed dispersive. In this situation, the different frequency components within a broadband radar pulse would travel at slightly different speeds, causing the pulse shape to change with time. However, the propagation characteristics of octave band radar signals in most earth materials remain largely unaffected by dispersion In many instances, the potential variation in the velocity of wave propagation over the frequency range of interest is small and can be ignored.

Depth Resolution. For traditional radar systems, it is accepted that two identical targets can be separated in range if they are 0.8 of a pulse width apart. In optics, Lord Rayleigh proposed that the resolving power of an instrument is when the principal intensity of one component coincides with the first intensity minimum of the other component. Many GPR pulses take the form of a Ricker wavelet (the second differential of a gaussian impulse), and an example of two pulses from targets is shown in Figure 21.5 where both the impulses and their envelopes are shown. When the targets are closer, as

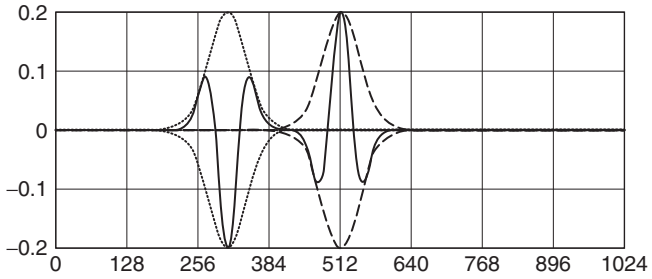


FIGURE 21.5 Two resolved Ricker wavelets

shown in Figure 21.6, although it is possible to distinguish their envelopes, it becomes increasingly difficult to resolve the actual pulses because the signal might be due to a resonance generated by a single target; hence, in the case of pulses with an envelope that has no minima, the 0.8 pulse width resolution criteria may not be optimal.

Essentially, range resolution is defined by the bandwidth of the received signal. A receiver bandwidth in excess of 500 MHz and typically 1 GHz is required to provide a typical resolution of between 5 and 20 cm, depending on the relative permittivity of the material.

When a number of features may be present, a signal having a larger bandwidth is required to be able to distinguish between the various targets and to show the detailed structure of a target. In this context, it is the bandwidth of the received signal that is important, rather than that of the transmitted wavelet. The material acts as a low pass filter, which modifies the transmitted spectrum in accordance with the electrical properties of the propagating medium. There are some applications of GPR, such as road layer thickness measurement, where the feature of interest is a single interface. Under such circumstances, it is possible to determine the depth sufficiently accurately by measuring the elapsed time between the leading edge of the received wavelet provided the propagation velocity is accurately known.

Although a greater depth resolution is achieved in wetter materials for a given transmitted bandwidth because of the reduced wavelength in high dielectric materials, earth materials with significant water content tend to have higher attenuation properties. This characteristic reduces the effective bandwidth, tending to balance out the change so that within certain bounds the resolution is approximately independent of loss within the propagating material.

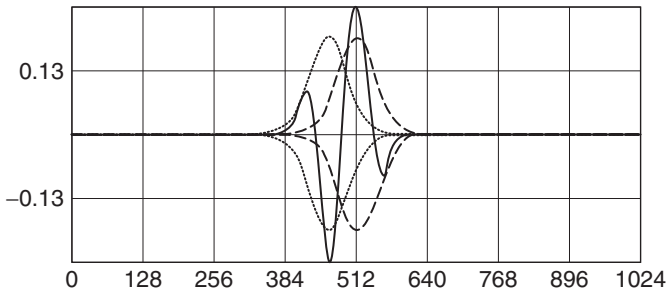


FIGURE 21.6 Two unresolved Ricker wavelets

Where interfaces are spaced more closely than one half wavelength, the reflected signal from one interface will become difficult to resolve with that from another.

It should be noted that the normal radar criteria for range resolution is less appropriate for the case of a weak target adjacent to a strong target, and there is no accepted definition of resolution for the case of unequal size targets.

Plan Resolution. The plan (*plan* is defined as a plane normal to the direction of propagation) resolution of a GPR system is important when localized targets are sought and when there is a need to distinguish between more than one at the same depth. Where the requirement is for location accuracy, which is primarily a topographic surveying function, the system requirement is less demanding.

The plan resolution is defined by the characteristics of the antenna and the signal processing employed. In general radar systems (apart from SAR), to achieve an acceptable plan resolution requires a high gain antenna. This necessitates a sufficiently large aperture at the lowest frequency to be transmitted. To achieve small antenna dimensions and high gain, therefore, requires the use of a high carrier frequency, which may not penetrate the material to sufficient depth. When selecting equipment for a particular application, it is necessary to compromise between plan resolution, size of antenna, the scope for signal processing, and the ability to penetrate the material. Plan resolution improves as attenuation increases, provided that there is sufficient signal to discriminate under the prevailing clutter conditions. In low attenuation media, the resolution obtained by the horizontal scanning technique is degraded, but only under these conditions do synthetic aperture techniques increase the plan resolution. Essentially the ground attenuation has the effect of placing a "window" across the SAR aperture, and the higher the attenuation the more severe the window. Hence, in high attenuation soils, SAR techniques may not provide any useful improvement to GPR systems. SAR techniques have been applied to GPR but very often in dry soils with low attenuation.

SAR techniques typically require measurements made using transmitter and receiver pairs at a number of antenna positions to generate a synthetic aperture or to focus the image. Unlike conventional radars, which generally use a single antenna, most GPR systems use separate transmit and receive antennas to provide receiver isolation. The GPR community refers to this as a bistatic mode, although actually the antenna system is closely spaced and mobile. This is different from the traditional radar community that associates the term *bistatic* with large separations.

21.3 MODELING

Models of the GPR situation range from a simple single frequency evaluation of path losses to complete 3D time-domain descriptions of the GPR and its environment. Modeling techniques include single frequency models, time-domain models, ray tracing, integral techniques, method of moments (MoM), and discrete element methods. The Finite-Difference Time-Domain (FDTD) technique has become one of the popular techniques and can be developed to run on most desktop computers with relative efficiency.

It should be noted that GPR systems often operate in intimate contact with the ground and very close to the target. Thus the antenna radiates in the near-field whereas some geophysical GPR systems operate at longer ranges (10 m to 2 km), and they

could be considered to operate in the Fresnel and even Fraunhofer (far-field) region. When the target is so close to the antenna, it interacts with the reactive fields of the antenna, and accurate models would reflect this mode of operation.

The most basic model uses the radar range equation and enables an estimate of received signal level, dynamic range, and probability of detection to be assessed. It has significant weaknesses in that most close-range GPR systems are operating in the near-field or even the reactive field of the antenna whereas the model assumes a far-field model. It is probably more relevant to the longer-range geophysical applications where the target is many tens of meters from the radar.

Many GPR receivers were originally based on sampling oscilloscope technology and the use of voltage became experimentally more useful. The most basic model for assessment of voltage signal level is derived from the radar range equation, which does have the limitations previously noted. However, it does enable a first order assessment of anticipated signal levels and an example is given in this section. The model is based on the equation for the voltage at the receiver as a function of range r and target radar cross-section σ and given by reference Rutledge and Muha.[30]

In the first model shown in Figure 21.7, the antenna is set at a height of 15 cm above the target (dielectric cylinders of 1 cm thickness, ranging in size from 0.05 m diameter to 0.5 m diameter). The target has a value of ε_r of 2.2 and the soil $\varepsilon_r = 5$ and $\tan\delta = 0.2$. The radiated pulse has a center frequency of 1 GHz and an output pulse peak voltage of 10 volts. The radar receiver has an equivalent bandwidth of 300 MHz to 3 GHz and an equivalent receiver noise voltage of $2.49.10^{-5}$ volts.

The probability of detection (PD) is derived from the error function of the signal-to-noise ratio, as shown in Figure 21.8. Note that these values only relate to the receiver noise and do not include external sources of false alarms due to clutter.

The most basic model is that of the transmission line equivalent and is useful for assessing the time-domain signature of a physical situation. A conceptually simple model can be used to gain an insight into the optimum center frequency of operation and is shown in Figure 21.9.

Each layer is modeled as equivalent impedance and the transmission and reflection coefficients are calculated for each interface. The velocity of propagation and the material losses are included although not the spreading losses. The reason for this is that the received A-scan would normally have time-varying gain applied in the receiver and signal processing, and to introduce spreading loss and then compensate is an inefficient modeling exercise. In the model only the first reflection is computed, although multiple internal reflections within each layer will be generated and a full representation should include these. The parameters of the layers are given in Table 21.3.

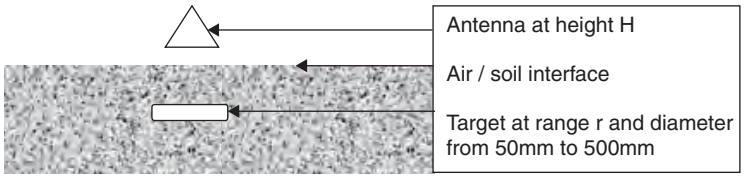


FIGURE 21.7 Physical layout of GPR system (*Courtesy IEE*)

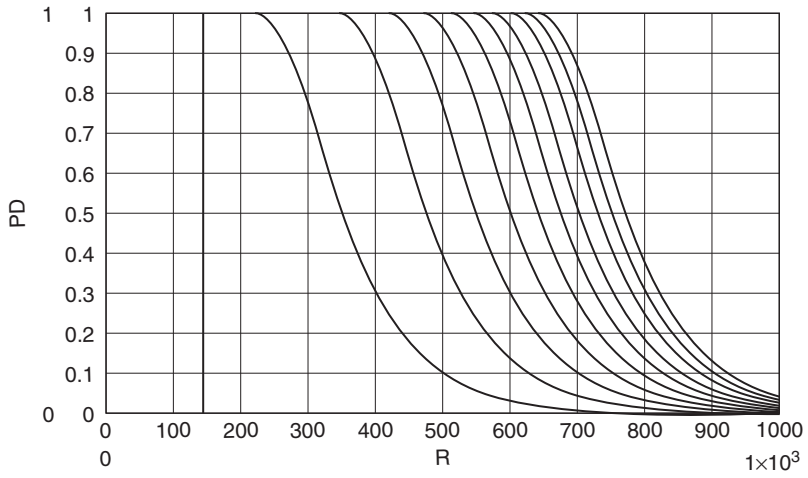


FIGURE 21.8 Graph of probability of detection as a function of target range in millimeters and target diameter (50 mm left-hand side to 500 mm right-hand side in increments of 50 mm). Ground surface is shown as a vertical line. (*Courtesy IEE*)

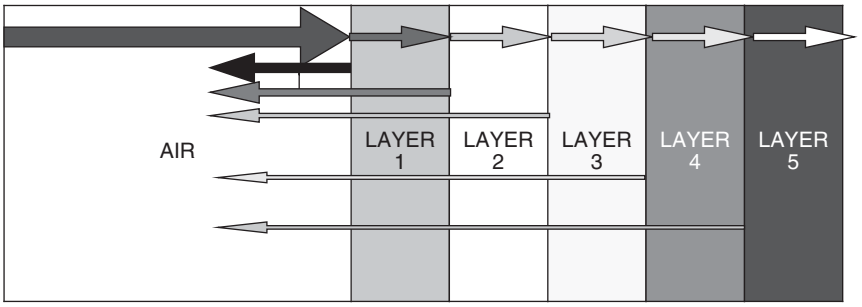


FIGURE 21.9 Layout of transmission line model (*courtesy IEE*)

TABLE 21.3 Layer Characteristics for Transmission Line Model

Layer	Range in Meters	Relative Dielectic Constant e_r	Loss Tangent	Material
0	0	1	0	Air
1	0.3	6	0.31	Lossy layer
2	0.6	1	0	Air void
3	0.85	9	0.01	Sub base
4	1	16	0.1	Wet base
5	infinite	25	0.1	Wet bedrock

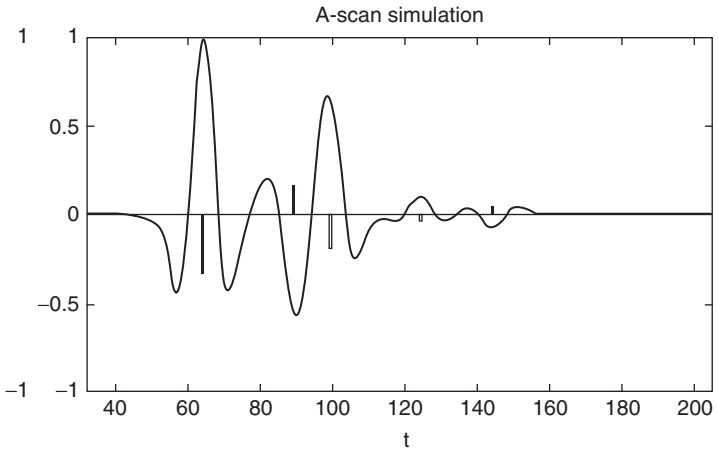


FIGURE 21.10 Simulation of A-scan using 300 MHz center frequency (*Courtesy IEE*)

The output from the model is shown in Figure 21.10 and Figure 21.11.

Finite-Difference Time-Domain (FDTD) methods can be used to model the field propagation of a typical GPR system. The antenna used for this purpose is a resistively loaded TEM horn, as described by Martel et al.[22] It is 35 cm long with an aperture of 10 cm by 30 cm. The TEM horn has ultrawideband capabilities from 200 MHz to 4 GHz. It is positioned above a metallic target buried in the ground as shown in Figure 21.12. The distance between the horn aperture and the air-ground, interface is 25 cm (different from the earlier model). The modeled target is a cylinder with a radius of 3.5 cm and a height of 5 cm. It is shallowly buried at about 2.5 cm below the air-ground interface.

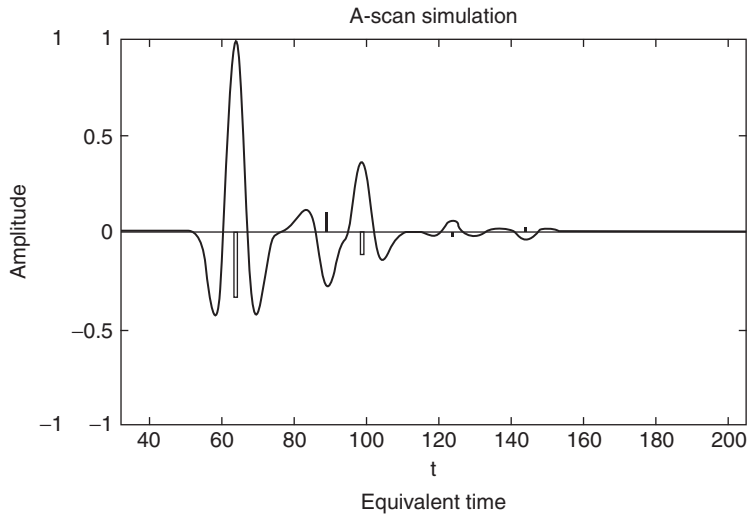


FIGURE 21.11 Simulation of A-scan using 500 MHz center frequency (*Courtesy IEE*)

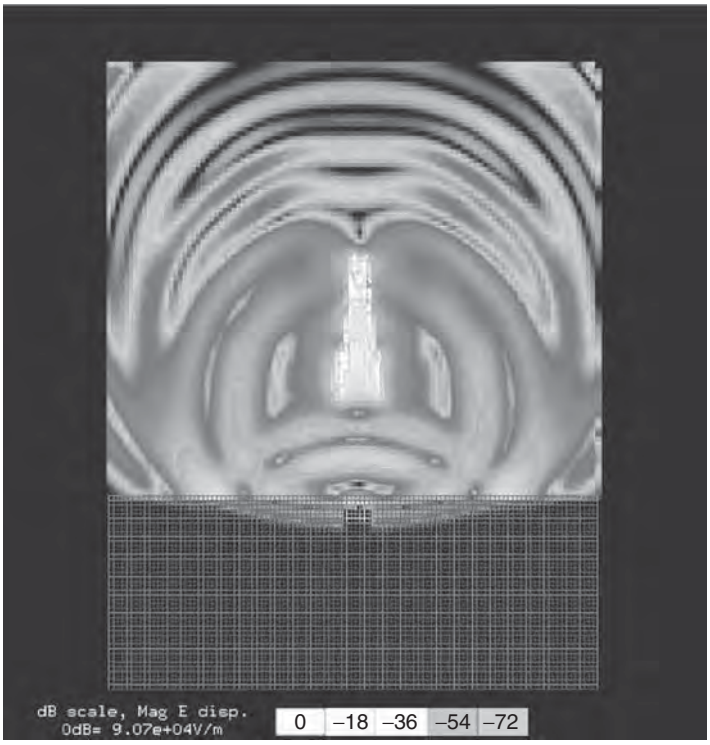


FIGURE 21.12 Electric field plot on a vertical cut plane after the main ground reflection (*Courtesy IEE*)

The ground is modeled as a uniform lossy material with a relative permittivity of 13 and a conductivity of 0.005 S/m (Siemens/meter). The air-ground interface is assumed to be perfectly flat.

From the field plot, one can recognize the antenna structure and the strong field region on and inside the horn plate. The buried object is also visible. The main reflection caused by the air-ground interface can clearly be seen coming back toward the antenna system. In addition, a weaker reflection coming from the buried object is starting to form and follows the air-ground interface reflection in time. This is a typical time-domain characteristic of standoff GPR system. Moreover, other physical phenomena can be observed such as the free space path loss and the reduction in velocity of propagation in the ground.

It should be noted that the process of physically scanning the antenna system over the target creates a hyperbolic image of the target, as shown in Figure 21.13. For the two-dimensional case (x = position on surface and z = depth to the target) of a material with known constant velocity, the measured time to the point reflector is given by t and then the distance to the point reflector is given by $z = vt/2$. At any position along the x-axis the distance z is also given by

$$z_i = \sqrt{(x_i - x_0)^2 + z_0^2} \tag{21.11}$$

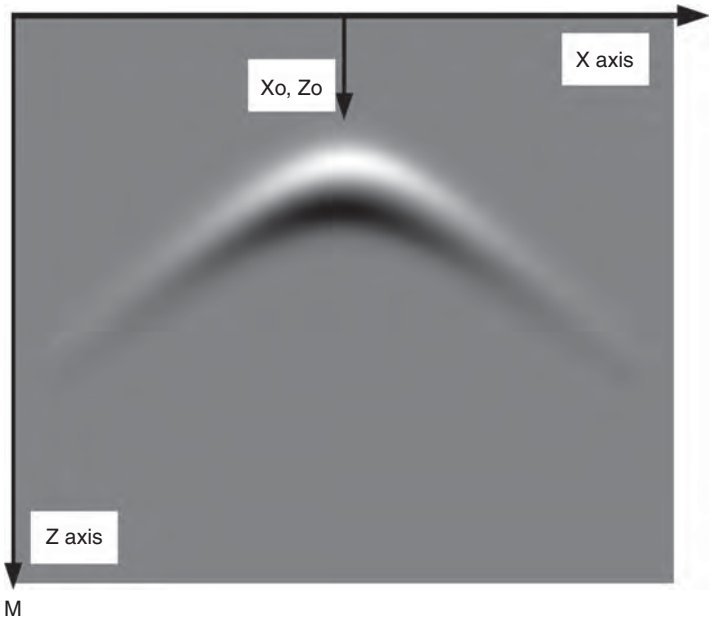


FIGURE 21.13 Typical hyperbolic image of GPR data from a reflector of circular cross section (*Courtesy IEE*)

This equation shows that the measured wavefront appears as a hyperbolic image or a curve of maximum convexity. Migration technique may be used to move or migrate a segment of an A-scan time sample to the apex of a curve of maximum convexity. The hyperbolic curve needs to be well-separated from other features and a good signal-to-noise ratio is needed for this technique to work well.

21.4 PROPERTIES OF MATERIALS

The determination of the dielectric properties of earth materials remains largely experimental. Rocks, soils, and concrete are complex materials composed of many different minerals in widely varying proportions, and their dielectric parameters may differ greatly even within materials that are nominally similar. Most earth materials contain moisture, usually with some measure of salinity. Since the relative permittivity of water is in the order of 80, even small amounts of moisture cause a significant increase of the relative permittivity of the material. A large number of workers have investigated the relationships between the physical, chemical, and mechanical properties of materials and their electrical and, in particular, microwave properties. In general, they have sought to develop suitable models to link the properties of the material to its electromagnetic parameters. Such models provide a basis for understanding the behavior of electromagnetic waves within these media. The real and imaginary dielectric losses

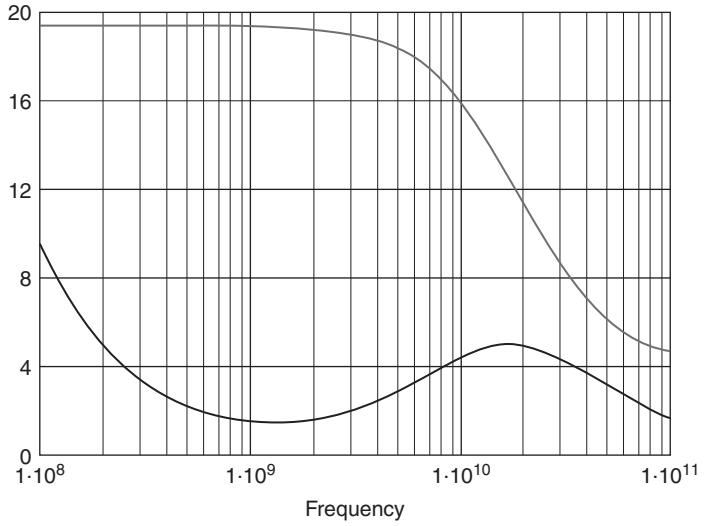


FIGURE 21.14 Dielectric properties ε' (upper) and ε'' (lower) of lossy soil as a function of frequency (*Courtesy IEE*)

as a function of frequency can be plotted over a wide frequency range and a typical result is shown in Figure 21.14.

Information on the geological properties of earth soils can be found in the Digital Soil Map of the World and Derived Soil Properties CD, published by the Food and Agriculture Organization of the United Nations. This enables the ten map sheets of the world to be classified in terms of parameters such as pH (concentration of hydrogen ions), organic carbon content, C/N (carbon-to-nitrogen) ratio, clay mineralogy, soil depth, soil moisture capacity, and soil drainage class. Such information is useful in assessing the potential of RF techniques and particularly GPR for particular geographic regions.

There are two benefits to understanding soil properties in relation to GPR. The first is to understand the applicability of GPR to particular soils and hence the possibility of using GPR to detect buried targets such as pipe, cables, landmines, etc. The second is to use GPR to characterize soils and soil properties.

GPR can provide a detailed map of the subsurface, which when combined with traditional soil survey methods can provide information on the type of soil, its extent laterally and in depth, the water table, the layering and features of the soil, and hence its local geology and history.

GPR has been used in the U.S. by the Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS) as a quality control tool for soil mapping and investigations. The use of GPR in soil survey activities has provided information about soil resources that would have been unobtainable by other means or would have been uneconomical to obtain. An example of the results of this work is shown in Figure 21.15.

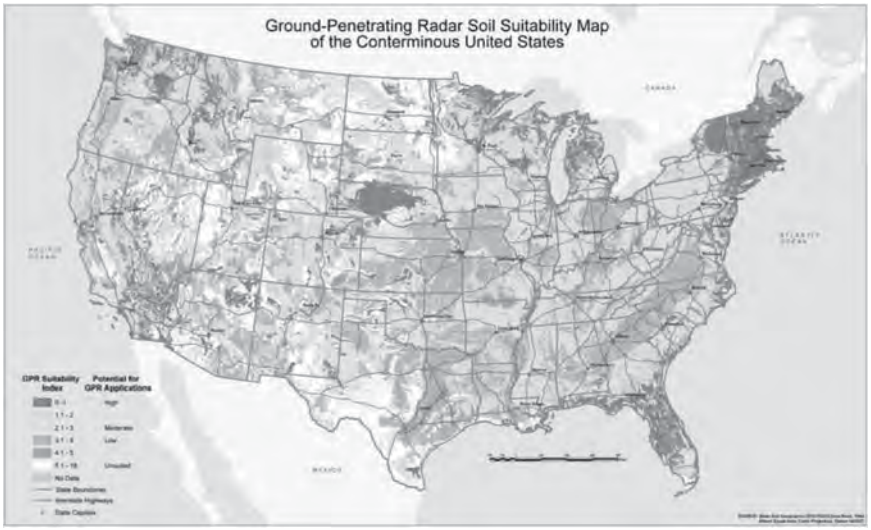

Ground-Penetrating Radar Soil Suitability Map
of the Conterminous United States
GPR Suitability Index
Potential for GPR Applications

FIGURE 21.15 GPR Soil Suitability Map of the Continental (Conterminous) United States (*Courtesy of USDA-NRCS*)

21.5 GPR SYSTEMS

The choice of system design is to a large extent governed by the type of target, the resolution required, and the anticipated ground attenuation and clutter. The depth range of the radar system is likely to be primarily defined by the soil attenuation, once a particular range of frequencies has been chosen. However, it can be shown that considerable variations (10–30 dB) in the sensitivity of competing system designs actually translate to relatively small changes in depth performance in lossy soils.

The selection of a suitable waveform for transmission, at least in terms of resolution, can be considered a function of the duration of the complex envelope of the signal. The output from most ultrawideband radar systems can be compared in terms of a time-domain representation of the waveform. Almost all types of radar can be assessed not just by their signal-to-noise and signal-to-clutter ratios but also by comparing their inherent range sensitivity. Such a procedure reveals the characteristics that control the radar performance. The design of a GPR system is defined by the modulation technique, and time domain, frequency domain, and pseudo-random-coded domain radar designs are most likely to be encountered. Frequency domain radars may use either stepped frequency or continuously swept frequency modulation. They transmit, on a repetitive basis, a nominally constant amplitude signal whose frequency increases in a linear progression from the lowest to the highest value.

Recovery of the receiver signal from noise may be achieved by either conventional bandpass filters or by the matched filter or Wiener filter.

The fundamental operation of a matched filter is correlation. The amplitude of each point in the output signal is a measure of how well the filter kernel matches the corresponding section of the input signal. The output of a matched filter does not necessarily look like the signal being detected, but if a matched filter is used, the shape of the target signal must already be known. The matched filter is optimal in the sense

that the peak signal output to mean noise ratio is greater than can be achieved with any other linear filter. This is not always the best filter to use for time-domain waveforms where the fidelity of the output may be a requirement.

The Wiener filter separates signals based on their frequency spectra. The gain of the Wiener filter at each frequency is determined by the relative amount of signal and noise at that frequency: The Wiener and matched filter must be carried out by *convolution,* making them extremely slow to execute.

The matched filter radar receiver provides an optimum linear processing of radar in the presence of noise. The radar signal is processed by a filter that cross-correlates the received waveform with a suitably time-delayed version of the transmitted waveform. The output results in an output in which the amplitude of the latter and its position in delay time is related to the target radar characteristic. This type of receiver is widely used to process chirp, step frequency, and pseudo-random-coded waveforms, and the design of such waveforms is extensively described in the literature.

Many commercial time-domain radar systems use a sampling receiver to downconvert the radar signals from the nanosecond time frame to a millisecond time frame that is easier to post process. However, a real disadvantage of the sampling receiver is its limited dynamic range due to the sampling diodes and inherently high noise level due to its wide bandwidth. Details of typical sampling receivers, which are essentially the same as sampling oscilloscopes, can be found in the literature, and once issues of sampling linearity in time are addressed, the generic design has formed the basis for the majority of commercial GPR systems.

A key parameter for most GPR systems is the mean power. The time-domain radar transmits, on a repetitive basis, a short duration impulse. Consequently, its peak power is significantly greater than its mean power. This is not the case with stepped frequency whose radiated power per spectral line is higher than the time-domain radar that gives an advantage in terms of transmitter peak signal capability compared with the impulse GPR.

21.6 MODULATION TECHNIQUES

There are three main modulation techniques: time domain, frequency domain, and pseudo-random-coded radar. GPR systems that transmit an impulse and receive the reflected signal from the target using a sampling receiver can be considered to operate in the time domain. GPR systems that transmit individual frequencies in a sequential manner and receive the reflected signal from the target using a frequency conversion receiver can be considered to operate in the frequency domain. The latter systems often reconstruct the downconverted frequencies to recover a time-domain replica of the signal.

All GPRs may have to detect signals from a target that may be –50 to –100 dB lower than the radiated signal at ranges in the order of a meter (6.6 ns in free space). In addition, the received signal will contain temporal scattering information on the target that can be exploited.

The temporal fidelity of received signal needs to be preserved and thus designers of GPR have to ensure that the receiver is not saturated by the transmitted signal, the antennas do not cause time sidelobes, and the receiver does not distort the received signal.

Time Domain. Most commercially available GPR systems use short pulses or impulses such as the Ricker wavelet, as shown previously in Figure 21.5. The high-speed sequential sampling approach used to acquire RF waveforms produces a low

SNR because the spectrum of the sampling pulse is a poor match for that of the received pulse. In general, the dynamic range of the sampling receiver is typically 60 dB, without range (time) varying gain. The effect of the range varying gain is to enable the lower amplitude signals from targets at a greater range to be amplified so as to be above the minimum sampling threshold signal level. This is equivalent to a linear receiver with a 90 dB or more dynamic range. Signal averaging or integration of the samples can increase the effective sensitivity by the amount of averaging and this can be typically 10 to 30 dB. The ratio of the peak transmitted signal to the mean receiver noise level can be up to 150 dB.

The antennas that can be used with time-domain GPR are limited to linear phase designs such as resistively loaded dipoles, TEM horns, or Impulse Radiating Antennas (IRAs). It should be noted that ultrawideband antennas fall into two classes, those that radiate a reasonably short impulse with low time sidelobes and fundamentally possess a linear phase-frequency characteristic. The alternative class of antennas, such as log periodics, have wideband frequency characteristics but nonlinear phase-frequency characteristics. Essentially, the latter class will cause the different frequency components of an impulse to be radiated at different times, hence dispersing the impulse. If such antennas are used with time-domain radar, the dispersive properties of the antenna used must be compensated by suitable post-processing filtering.

The time-domain radar system transmits a sequence of pulses, typically of amplitudes within the range between 20 V to 200 V and pulse widths within the range between 200 ps to 50 ns at a pulse repetition interval of between several hundred microseconds to one microsecond, depending on the system design. The impulse generator is generally based on the technique of rapid discharge of the stored energy in a short transmission line. The most common method of achieving this is by means of a transistor operated in avalanche breakdown mode used as the fast switch and a very short length of transmission line. It is quite feasible to generate pulses of several hundred kV albeit at long repetition intervals. The output from the receive antenna is applied to a flash A/D converter or a sequential sampling receiver. The latter normally consists of an ultrahigh speed sample and hold circuit. The control signal to the sample and hold circuit, which determines the instant of sample time, is sequentially incremented at each pulse repetition interval. For example, a sampling increment of $t = 100$ ps is added to the previous pulse repetition sampling interval to enable sampling of the received signal at regular intervals as indicated in Figure 21.16.

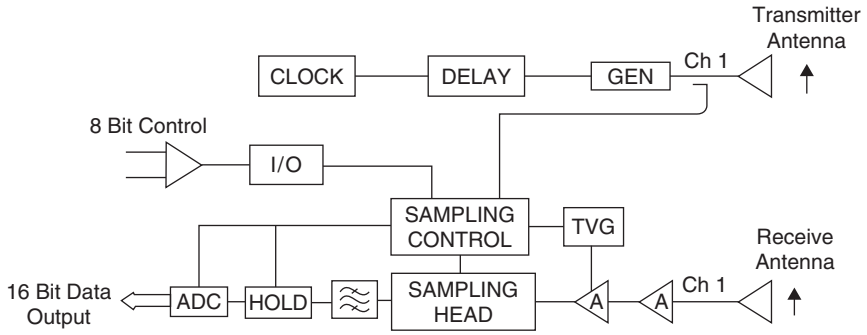


FIGURE 21.16 Typical sampling receiver used for time-domain GPR systems

The principle of the sampling receiver is, therefore, a downconversion of the radio frequency signal in the nanosecond time region to an equivalent version in the micro- or millisecond time region. The incrementation of the sampling interval is terminated at a stage when, for example, 256, 512, or 1024 sequential samples have been gathered. The process is then repeated. There are several methods of averaging or "stacking" the data; either a complete set of samples can be gathered and stored and further sets added to the stored data set, or alternatively, the sampling interval is held constant for a predetermined time to accumulate and average a given number of individual samples. The first method needs a digital store but has the advantage that each waveform set suffers little distortion if the radar is moving over the ground.

The second method does not need a digital store and a simple low-pass analogue filter can be used. However, depending on the number of samples that have been averaged, the overall waveform set can result in being "smeared" spatially if the radar is moving at any speed. The stability of the timing increment is very important and generally this should be 10% of the sampling increment; however, practically a stability in the order of 10 ps to 50 ps is achieved. The effect of timing instability is to cause a distortion, which is related to the rate of change of the RF waveform. Evidently, where the RF waveform is changing rapidly, jitter in the sampling circuits results in a very noisy reconstructed waveform. Where the rate of change of signal is slow, jitter is less noticeable. Normally, control of the sampling converter is derived from a sample of the output from the pulse generator to ensure that variations in the timing of the latter are compensated automatically. The key elements of this type of radar system are the impulse generator, the timing control circuits, the sampling detector, and the peak hold and analogue to digital converter.

Frequency Domain Radar. The main potential advantages of the frequency domain radar are the wider dynamic range, lower noise figure, and higher mean powers that can be radiated. There are two main types of frequency domain radar, Frequency Modulated Carrier Wave (FMCW) and Stepped Frequency Carrier Wave (SFCW). FMCW radar transmits a continuously changing frequency over a chosen frequency range on a repetitive basis. The received signal is mixed with a sample of the transmitted waveform and results in a difference frequency, which, although fundamentally related to the phase of the received signal, is a measure of its time delay and hence range of the target. The difference frequency or intermediate frequency (IF) must be derived from an I/Q mixer pair if the information equivalent to a time-domain representation is required (i.e., to reconstitute an impulse), as a single-ended mixer only provides the modulus of the time-domain waveform. The basic FMCW radar system is particularly sensitive to certain parameters. In particular, it requires a high degree of linearity of frequency sweep with time to avoid spectral widening of the IF and hence degradation of system resolution. Dennis and Gibbs[23] made an assessment of the sensitivity of time sidelobe level to linearity and showed the ratio of sidelobe to peak level was dependent on the sweep linearity. Practically, the effect of a nonlinearity of a few percent is to cause significant time sidelobes, as this needs to be compensated in the transmitter modulator design.

The SFCW radar transmits a series of incremental frequencies and stores the received IF signal to then carry out a Fourier transform reconstruction of the time-domain equivalent waveform. The SFCW has found many applications in GPR because the requirements on scan rate are relatively modest. The impact of mobile communications technology has had a significant impact on reducing the cost of radar components for this design. Two forms of the synthesized radar can be considered.

The first and simplest system is stepped frequency continuous wave radar. The second form is more complex in that each individual frequency is appropriately weighted in amplitude and phase prior to transmission. Normally, the radar is calibrated both to establish a reference plane for measurement as well as to reduce the effect of variations in the frequency characteristics of components and antennas.

A much wider class of antenna is available for use by the designer of frequency domain radars. The noise floor of the receiver is much lower than the time-domain equivalent, simply by virtue of its lower bandwidth and hence lower thermal noise. Typically, a sensitivity of –120 dBm is found and a system peak transmitted signal to mean receiver noise range of 180 dB is feasible. It should be noted that the IF bandwidth of the receiver in FMCW and SFCW systems can be made relatively small whereas the sampling receiver in the time-domain receiver has a bandwidth of many GHz and hence a poor noise performance.

The main potential advantage of a stepped frequency or FMCW GPR is its ability to adjust the range of frequencies of operation to suit the material and targets and electromagnetic environment under investigation if the antenna has an adequate passband of frequencies. It can radiate a higher mean power level per spectral line than the time-domain radar, and its ability to integrate the received signal level improves the system sensitivity. The calibration of the radar does, of course, depend on stable system characteristics and antenna parameters that are invariant with the spacing of the front surface and the antenna. Although on first consideration, frequency domain radars should offer a superior sensitivity to time-domain radars, because of their lower IF receiver bandwidth and hence thermal noise, both the type of receiver and the range sidelobes of the radiated spectrum may result in an equivalent or worse sensitivity in terms of range resolution as discussed above.

Pseudo-random-coded Radar. Work has been carried out on pseudo-random-coded modulation techniques for GPR. The main advantage of this method is that the energy transmitted is spread more evenly over the spectrum than with any other modulation method and hence the likelihood of interference to other users of the spectrum is minimized. In addition, the chances of other users of, say, mobile phones interfering with the GPR operator are also reduced. The mean power is the lowest of any of the modulation schemes and this is helpful in meeting regulatory requirements.

The transmitted signal has noise-like characteristics and the received signal is cross-correlated with a sample of the transmitted signal. The range of the target is given by the time position of the cross-correlated signal and the amplitude by the peak of the cross-correlated signal. Control of the cross-correlation sidelobes is vital to achieve good range resolution and the sidelobes are affected by the antenna and system characteristics as well as the duration and randomness of the transmitted waveform. Further information is given by Narayanan[24] and Sachs et al.[25,26]

21.7 ANTENNAS

In the ultrawideband case, the radar antennas are considered in terms of their transfer function rather than their gains or effective apertures. In many cases, a separate transmit and receive antenna is used; hence their transfer functions may not be identical. The type of antenna that is used with ultrawideband radar has an important role in defining the performance of the radar.

Element antennas are characterized by linear polarization, low directivity, and relatively limited bandwidth, unless either end-loading or distributed-loading techniques are employed in which case bandwidth is increased at the expense of radiation efficiency. An elemental antenna is a dipole in contrast to aperture antennas such as horns. A normal short dipole antenna fed with a very short current impulse will radiate from the feed points and the end of the element because of the latter's discontinuity as far as current flow is concerned. The current impulse will be reflected from the end of the dipole and travel up and down the dipole causing a series of impulses of radiation. This extends the time signature of the radiated waveform and degrades the range resolution of the system. This effect is shown in Figure 21.17 where the radiation pattern is shown at a time several pulse durations after application. The outer perimeter represents the energy radiated at time zero followed at intervals by the radiation from the feed points and the end of the element.

As it is required to radiate only a very short impulse, it is important to eliminate the reflection discontinuities from the feed points and ends of the antenna either by end loading or by reducing the amplitude of the charge and current reaching the ends. The latter can be achieved either by resistively coating the antenna or by constructing the antenna from a material such as Nichrome, which has a defined loss per unit area. In this case, the antenna radiates in a completely different way as the applied charge becomes spread over the entire element length, and hence, the centers of radiation are distributed along the length of the antenna. Typical radiated field patterns for a resistively loaded dipole are shown in Figure 21.18. However, the reduction in impulse duration is achieved at the expense of efficiency and loaded antennas may have efficiencies as low as 10%.

The types of antennas that are useful to the designer of ultrawideband radar fall into two groups: dispersive antennas and nondispersive antennas. Dispersive antennas have a nonlinear phase/frequency response whereas nondispersive antennas have a substantially linear phase/frequency response. Examples of dispersive antennas that have been used in ultrawideband radar are the exponential spiral, the Archimedean

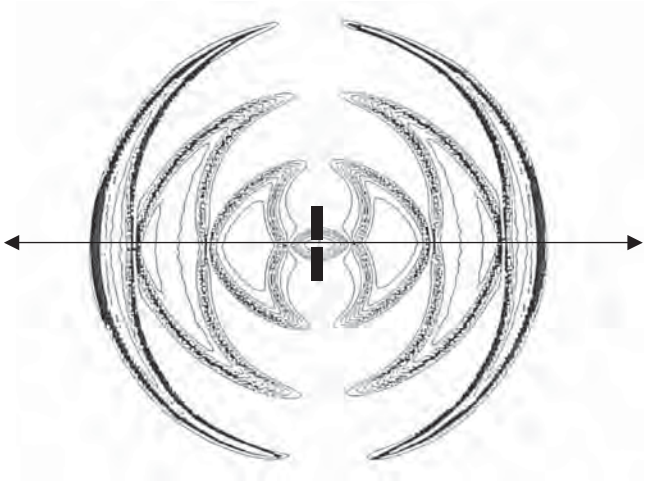


FIGURE 21.17 Radiated field pattern from a conducting dipole element due to an applied impulse (*Courtesy IEE*)

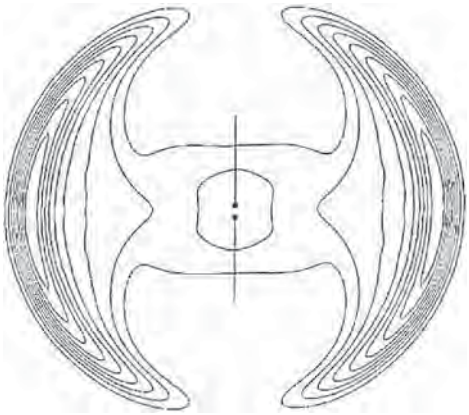


FIGURE 21.18 Radiated field pattern from a resistively loaded dipole element due to an applied impulse (*Courtesy IEE*)

spiral, the logarithmic planar antenna, the Vivaldi antenna, slot antennas, and the exponential horn. The impulse response of this class of antennas generally results in a waveform whose time frequency response is extended (and is similar to a chirp, albeit with an inconstant amplitude) if the input is an impulse.

Examples of nondispersive antennas are the TEM horn, the bicone; the bow tie; the resistive, lumped element-loaded antenna, and the continuously, resistively loaded antenna. The input voltage driving function to the terminals of the antenna in impulse radar is typically a narrow gaussian pulse of 200 ps, and this requires the impulse response of the antenna to be extremely short. The main reason for requiring the impulse response to be short is that it is important that the antenna does not distort the input function and generate time sidelobes. These time sidelobes would obscure targets that are close in range to the target of interest; in other words, the resolution of the radar can become degraded if the impulse response of the antenna is significantly extended.

However, in principle, all antennas are dispersive to some extent but nondispersive antennas do not need correction in the signal processing, which reduces the overall complexity of the radar processing. The very short range operation of many GPR systems enables operation of antennas in a way that does not conform to traditional analytic models of antenna gain and aperture.

Horn antennas have found most use with FMCW ultrawideband radars where the generally higher frequency of operation and relaxation of the requirement for linear phase response permit the consideration of this class of antenna. FMCW ultrawideband radars have used an offset paraboloid fed by a ridged horn. This arrangement was designed to focus the radiation into the ground at a slant angle to reduce the level of the reflection from the ground. Care needs to be taken in such arrangements to minimize the effect of back and sidelobes from the feed antenna, which can easily generate reflection from the ground surface.

One method of radiating circular polarization is to use an equi-angular spiral antenna. The dispersive nature of this type of antenna causes an increase in the duration of the transmitted waveforms, and the radiated pulse takes the form of a “chirp”

in which high frequencies are radiated first, followed by the low frequencies. A "spiking" filter, which may take the form of a conventional matched filter, however, may compensate for this effect, or a more sophisticated filter such as Wiener filter, which recovers the original shape of the waveform, applied to the antenna.

It is important to appreciate the effect of the material in close proximity to the antenna. In general this material, which in most cases will be soil or rocks or indeed ice, can be regarded as a lossy dielectric and by its consequent loading effect can play a significant role in determining the low frequency performance of the antenna and hence GPR. The behavior of the antenna is intimately linked with the material and, in the case of borehole radars, the antenna actually radiates within a lossy dielectric, whereas in the case of the GPR working above the surface, the antenna will radiate from air into a very small section of air and then into a lossy half space formed by the material. The behavior of antennas both within lossy dielectrics and over lossy dielectrics is well reported. The propagation of electromagnetic pulses in a homogeneous conducting earth has been modeled by Wait[27] and King,[28] and the dispersion of rectangular source pulses suggests that the time-domain characteristics of the received pulse could be used as an indication of distance.

The interaction between the antenna and the lossy dielectric half space is also significant as this may cause modification of the antenna radiation characteristics both spatially and temporally and should also be taken into account in the system design. In the case of an antenna placed on an interface, the two most important parameters are the current distribution and the radiation pattern. At the interface, currents on the antenna propagate at a velocity, which is intermediate between that in free space and that in the dielectric. In general, the velocity is retarded by the $\sqrt{(\varepsilon_r + 1)/2}$. The net result is that evanescent waves are excited in air, whereas in the dielectric the energy is concentrated and preferentially induced by a factor of n^3:1.

The respective calculated far-field power density patterns, in both air and dielectric, are given by Rutledge[29] (see Table 21.4), and these are plotted for a relative dielectric constant of 9 in Figure 21.19 and Figure 21.20. For comparison, the far-field pattern of a dipole radiating into free space is shown in Figure 21.21.

The above expressions in Table 21.4 assume that the current source contacts the dielectric, whereas a more general condition is when the antenna is just above the dielectric. A significant practical problem for many applications is the need to maintain sufficient spacing to avoid mechanical damage to the antenna. It can, therefore, be appreciated that the effect of changes in distance between the antenna and half space is to cause significant variation in the resultant radiation patterns in the dielectric.

A particularly useful antenna capable of supporting a forward traveling TEM wave is the TEM horn. In general, such antennas consist of a pair of conductors either

TABLE 21.4 Power Density Patterns in Air and Dielectric

	Power	Power
Plane	Radiation pattern in dielectric x direction	Radiation pattern in dielectric y direction
H	$\alpha\left(\frac{\cos\theta_a}{\cos\theta_a + \eta\cos\theta_d}\right)^2$	$\alpha\eta\left(\frac{\eta\cos\theta_d}{\cos\theta_a + \eta\cos\theta_d}\right)^2$
E	$\alpha\left(\frac{\cos\theta_a\cos\theta_d}{\eta\cos\theta_a + \cos\theta_d}\right)^2$	$\alpha\eta\left(\frac{\eta\cos\theta_a\cos\theta_d}{\eta\cos\theta_a + \cos\theta_d}\right)^2$

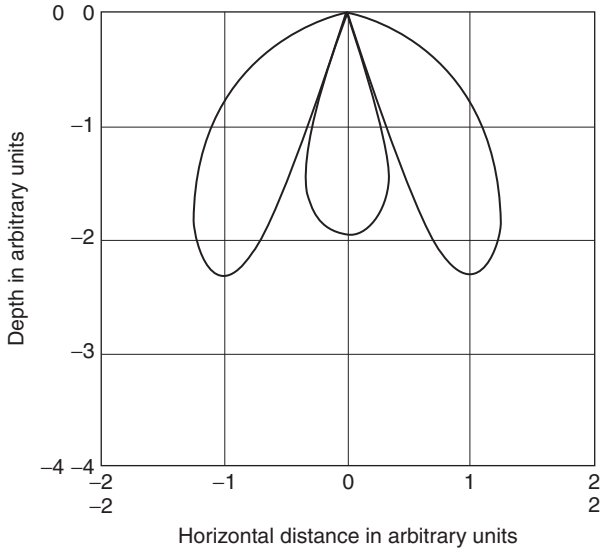


FIGURE 21.19 E-plane plot of far-field power density of a current element radiating into a lossless material of dielectric constant of 9

flat, cylindrical, or conical in cross section, forming a V structure in which radiation propagates along the axis of the V structure. Although resistive termination is used, this type of antenna has a directivity on the order of 10–15 dB depending on size; hence, useful gain can still be obtained even with a terminating loss on the order of

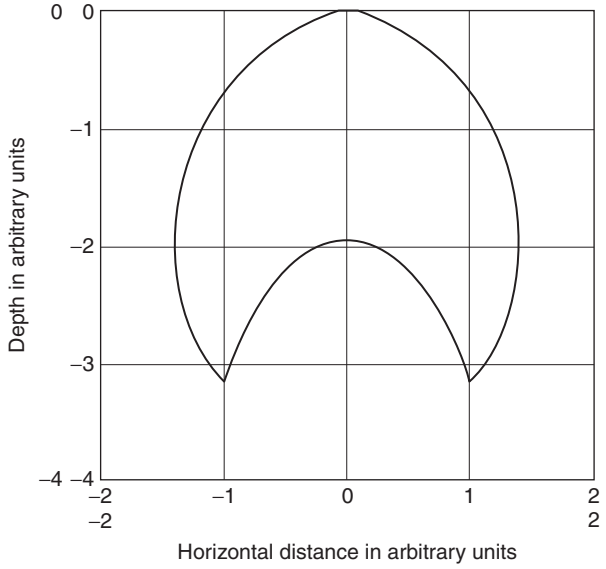


FIGURE 21.20 H-plane plot of far-field power density of a current element radiating into a lossless material of dielectric constant of 9

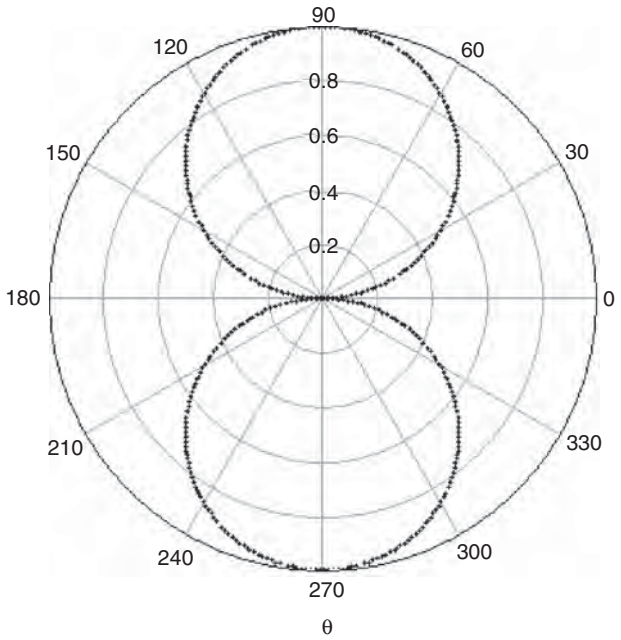


FIGURE 21.21 H-plane plot of far-field power density of a current element radiating into free space

3 dB to 5 dB. A further development of the TEM horn is given by Martel[21] in which the antenna is composed of a set of spread "fingers" forming the shape of a horn as shown in Figure 21.22. Each finger is a wire with a diameter of 1 mm and is resistively loaded at different locations along the length of the antenna. The feed component for the antenna is comprised of a 50 Ohm coax, feeding a tapered parallel plate waveguide with a width of 30 mm and a height of 7 mm. A taper along the width of the top parallel plate is used as a transition to transform a 50 Ohm unbalanced line into a 50 ohm balanced line.

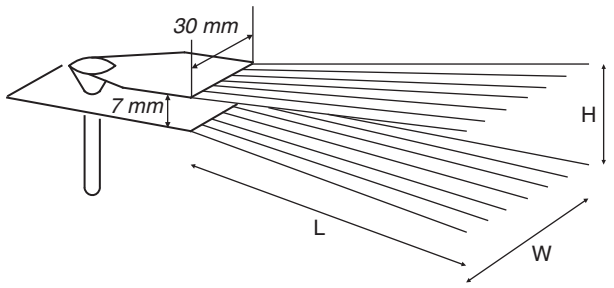


FIGURE 21.22 Antenna and feed geometry of loaded TEM horn, $L = 30$ cm (*Courtesy IEE*)

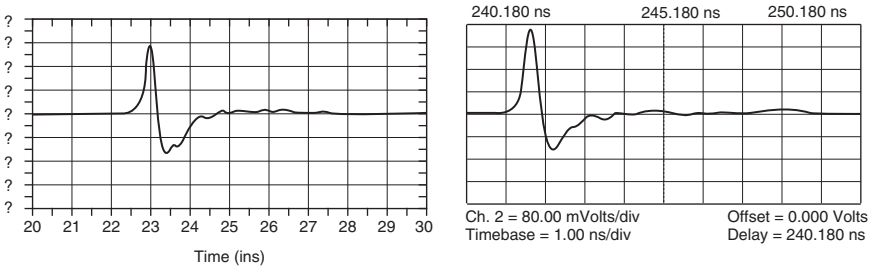


FIGURE 21.23 Predicted and actual pulse responses of loaded TEM horn antenna (*Courtesy IEE*)

The predicted and actual time-domain pulse of the optimum design is shown in Figure 21.23.The shape of the time-domain antenna response is similar to a second derivative gaussian signal. It can be seen that most of the internal reflections have been suppressed. The rate of decrease for the unwanted ringing is better than 9 dB/ns. The VSWR was better than 2:1 from 100 MHz to 5.8 GHz.

There are many configurations of antenna that can be used; crossed dipoles and parallel dipoles are the most popular. The main reason for the use of two antennas is that TR switches that are fast enough for GPR are not yet available.

21.8 SIGNAL AND IMAGE PROCESSING

The three basic permutations of GPR data presentation are shown in Figure 21.24. The most basic GPR data record is an A-scan. An A-scan provides amplitude-time record of a single measurement over a target. Only amplitude-range information is plotted. GPR is generally used in such a way as to generate a sequence of A-scans related to the survey position on the ground surface. This sequence can be termed a B-scan and an example is shown in Figure 21.24. This effectively represents one axis (z) depth and the orthogonal axis (x or y) linear position. The amplitude of the signal may be shown as a series of overlapping signals or, alternatively, a "wiggle plot" (borrowed from seismic terminology) or a grayscale-coded intensity plot or a pseudo-color image. In the modeled example shown, the hyperbolic spreading of the target spatial response can be seen. As shown in Figure 21.24, a C-scan consists of a plan view (x, y plane over a defined range of depth z). Note that these terms are not the same as used in conventional radar display.

The received time waveform can be described as the convolution of a number of time functions each representing the impulse response of some component of the radar system in addition to noise contributions from various sources. Note that two antennas are used: one transmits and one receives.

$$s_r(t) = s_s(t) \otimes s_{\text{af}}(t) \otimes s_c(t) \otimes s_{\text{gf}}(t) \otimes s_t t) \otimes s_{\text{gr}}(t) \oplus s_{\text{ar}}(t) + n(t) \qquad (21.12)$$

where

$s_s(t)$ = signal applied to the antenna
$s_{\text{ad}}(t)$ = antenna impulse response
$s_c(t)$ = antenna cross coupling response

$s_{gd}(t)$ = ground impulse response
$s_t(t)$ = impulse response of target
$n(t)$ = noise

(d denotes direction—f being forward and r reverse direction)

Each contribution has its own particular characteristics that need to be considered carefully before application of a particular processing scheme. Ideally the signal applied to the antenna should be a Dirac function but practically it is more like a skewed gaussian impulse of defined time duration. Most antennas used in surface penetrating applications have a limited low frequency response and tend to act as high-pass filters, effectively differentiating the applied impulse and hence creating a time-limited function. In most cases, near identical antenna are used, and if these are spaced sufficiently far from the ground surface, then $s_{af}(t) = s_{ar}(t)$. In the case of antennas operated in close proximity to the ground, then both $s_{af}(t)$ and $s_{ar}(t)$ are variant with

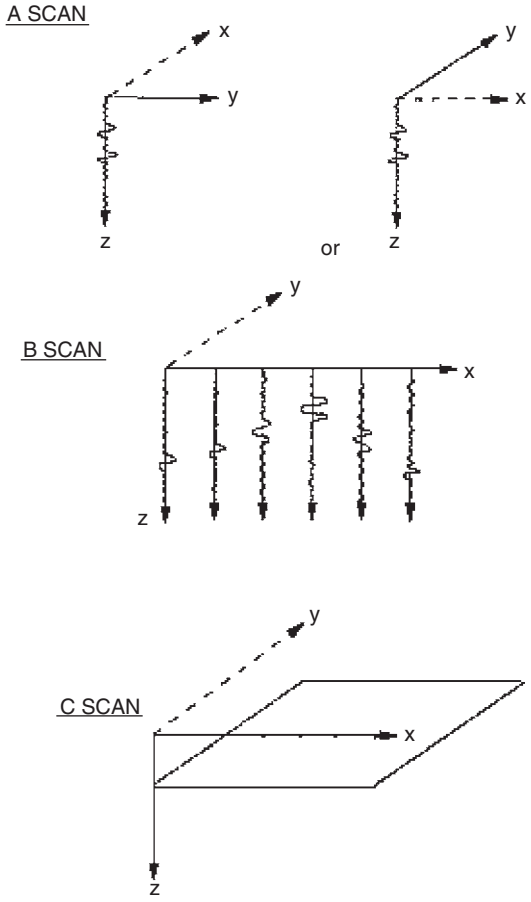


FIGURE 21.24 Coordinate system for scan description (*Courtesy IEE*)

changes in the ground surface electrical parameters. Essentially the impedance of the antenna is changed by its proximity to the ground; hence it cannot be considered to have a stable impulse response.

Any processing scheme that relies on invariant antenna parameters should take into account the mode of operation of the antennas and the degree of stability that is practically realizable. The antenna cross coupling response $s_c(t)$ is composed of a fixed contribution $s'_c(t)$ due to antenna cross coupling in air and a variable contribution $s''_c(t)$ due to the effect of the ground or nearby objects. Hence $s_c(t) = s'_c(t) + s''_c(t)$. It has been found possible to reduce the amplitude of $s_c(t)$ to very low levels, in the case of crossed dipole antennas to below –70 dB simply by attention to the precision of construction, and in the case of parallel dipole antennas to below –60 dB by the interposing between the antennas of suitable absorbing material. However, $s''_c(t)$ can be significantly larger and degrades the overall value of $s_c(t)$ to –40 dB. The value of $s_a(t)$ is determined by any local inhomogeneities in the soil or by any covering material whether of mineral or vegetable origin. There is unfortunately little that can be done to predict variations in $s''_c(t)$, and it is not amenable to treatment by many processing algorithms. The variation in $s''_c(t)$ is much greater with the crossed dipole antenna than the parallel dipole. The ground impulse response $s_g(t)$ can be determined from its attenuation and dielectric constant across the frequency range of interest.

The target impulse response can be composed of the convolution of the wanted target response, together with many other reflectors, which may not be wanted by the user but which are valid reflecting targets as far as electromagnetic waves are concerned. The time separation of the targets is related to their physical spacing as well as the velocity of propagation, which can vary depending on the material properties.

Where the targets are well separated in range, it is relatively straightforward to separate the radar reflections, but this becomes progressively more difficult as targets become closer together, as instead of separable time-domain signatures the reflections will merge together.

As the antennas generally used for GPR have poor directivity, the pattern of the reflected waveform in the B-scan represented the spatial convolution of the antenna pattern with the target. The reader is referred to Figure 21.13, shown previously, which illustrates this effect. This spatial pattern does not represent an image of the object of interest and much effort has been made to develop methods to reconstruct the target image from GPR data.

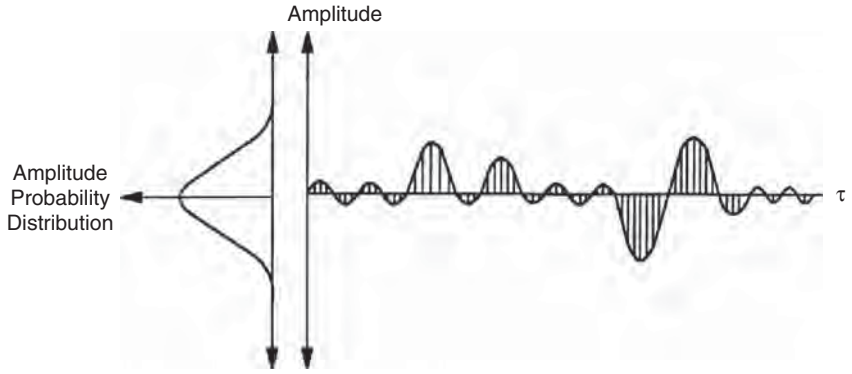


FIGURE 21.25 Envelope of A-scan sample time series (*Courtesy IEE*)

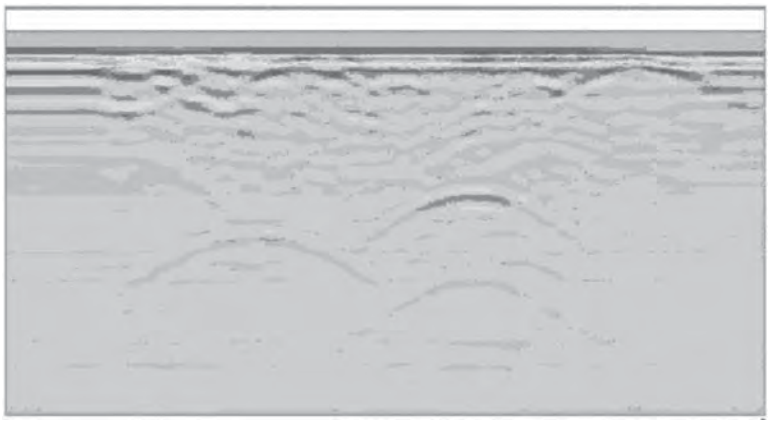


FIGURE 21.26 B-scan unprocessed data (*Courtesy IEE*)

Deconvolving the image using any of the following processes can do this; synthetic aperture processing, conjugate gradient methods, and reverse time migration are extensively reported in the literature. Many of these techniques work well on isolated targets such as pipes, which have well-defined geometrical boundaries. The situation is more difficult with stratified layers and, of course, anisotropic materials.

When a reconstructed image of the buried object is created, whether as a B-scan or C-scan (area at a particular range of depths), it is necessary to interpret the radar image as being generated by a physical structure. This is not always easy in the case of a cluttered image, and a great deal still depends on the field experience of the operator. Examples of unprocessed B-scan data and the same corrected for spreading loss and attenuation are shown in Figure 21.26 and Figure 21.27.

Although a C-scan is essentially an x, y plane at a selected value of Z or range of values of Z, many of the processes described in the previous section can be applied.

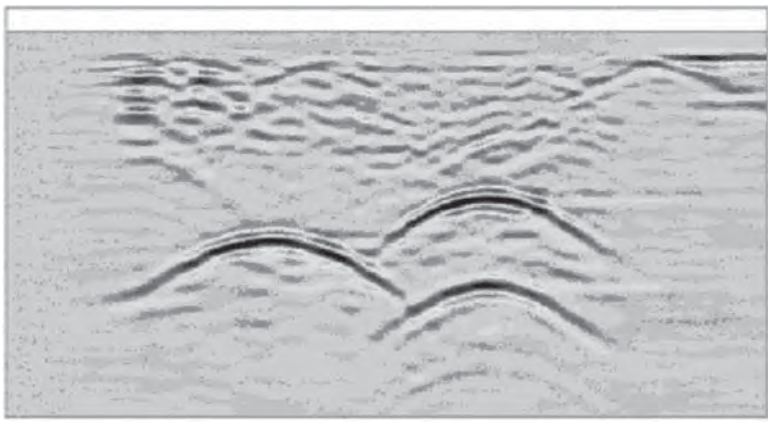


FIGURE 21.27 B-scan data corrected for spreading loss and attenuation (*Courtesy IEE*)

Remembering that a GPR B-scan results in a hyperbolic cross section from a target, then an area scan (C-scan) will result in a hyperboloid of revolution whose vertical axis runs through the target. A plane orthogonal to the vertical axis will generate a circular feature whose radius increases with depth. A typical example is given in Figure 21.28, which shows the C-scans from an anti-tank mine buried at different depths with the center of the mine shown as an overlay. These images represent an unfocussed representation of the target as a result of the 3D spatial convolution of the antenna pattern with the target.

The variability of ground conditions, as well as the physics of EM wave propagation and reflection, must be carefully taken into account in attempting to classify targets even after image processing. For example, the depth image of a void is always apparently smaller than its physical size; corner reflectors of any reasonable size generate large apparently discontinuous reflection images; and conductive targets, which reverberate by means of stored energy, create extended depth images. The image of a buried target generated by a GPR will not, of course, correspond to its geometrical representation. The fundamental reasons for this are related to the ratio of the wavelength of the radiation and the physical dimensions of the target. In most cases for GPR, the ratio is close to unity. This compares very differently with an optical image, which is obtained with wavelengths such that the ratio is considerably greater than unity (typically 100,000:1). In GPR applications, the effect of combinations of scattering planes, for example, the corner reflector, can cause "bright spots" in the image, and variations in the velocity of propagation can cause dilation of the aspect ratio of the image. Although many images can be focused to reduce the effect of antenna beam spreading, regeneration of a geometric model is a much more complex procedure and is not usually attempted.

The general objective of signal processing as applied to GPR is to present either an image that can readily be interpreted by the operator or to classify the target return with respect to a known test procedure or template.

The general processing problem encountered in dealing with GPR data is in the widest sense the extraction of a localized wavelet function from a time series, which displays very similar time-domain characteristics to the wavelet. This time series is generated by signals from the ground and other reflecting surface, as well as internally from the radar system. Unlike conventional radar systems in which the target can generally be regarded as being in motion compared with the clutter, in the GPR case, the target and the clutter are spatially fixed and the radar antenna is moved with respect to the environment.

It is assumed that data are recorded to an adequate resolution. Most antennas used in surface penetrating applications have a limited low-frequency response and tend to act as high-pass filters effectively differentiating the applied impulse. As the ground acts as a low pass filter, the ground largely defines the bandwidth of reflected signal.

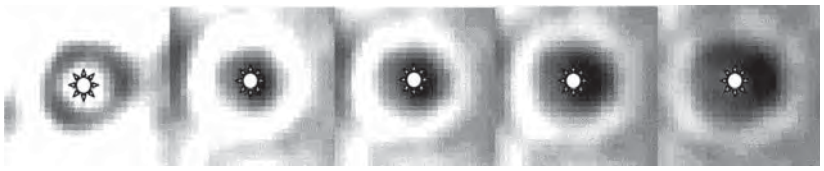


FIGURE 21.28 Sequence of unfocussed C-scans of a set of a buried AT mine targets at depth increments of 10 mm (*Courtesy IEE*)

In the case of antennas operated in close proximity to the ground, the antenna characteristics may vary as a result of changes in the ground surface electrical parameters. Any processing scheme that assumes the antenna parameters remain constant needs to account for the mode of operation of the antennas and the degree of stability that is practically realizable. This is a particular issue for GPR and needs careful attention to reduce the effect of antenna-ground surface interaction.

Some of the ancillary requirements of an operational GPR system need to be considered. Accurate, high-resolution, low-cost position referencing systems for use with radar for subsurface survey techniques are now available. It is important that data can be related to a true geographic reference particularly when filed on digital mapping systems and used to define areas of safe working.

Another consideration is the plane of polarization of the electromagnetic energy. For targets with one large area dimension such as a pipe, the radar cross-scattering section will be larger when the polarization vector is in line with the pipe. This means that any area that is surveyed with, say, parallel dipoles must be surveyed in orthogonal directions to ensure that no targets are missed. The same requirement also relates to crossed dipole antennas.

21.9 APPLICATIONS

It is only possible to provide a brief summary of the wide variety of the applications for GPR, which has in some cases become an established and routine method of subsurface investigation. GPR, in the hands of an expert, provides a safe and noninvasive method of conducting speculative searches without the need for unnecessary disruption and excavation. A typical example of a GPR image is shown in Figure 21.29 with the various features identified.

GPR has significantly improved the efficiency of the exploratory work that is fundamental to the construction and civil engineering industries, the police and forensic sectors, security/intelligence forces, and archaeological surveys.

GPR has been very successfully used in forensic investigations. The most notorious case being in the United Kingdom in 1994, when the gravesites, under concrete and in the house of Fred West, of the victims of the serial murderer were pinpointed. In Belgium, the gravesites of the victims of the pedophile, Duteous, were detected in 1996.

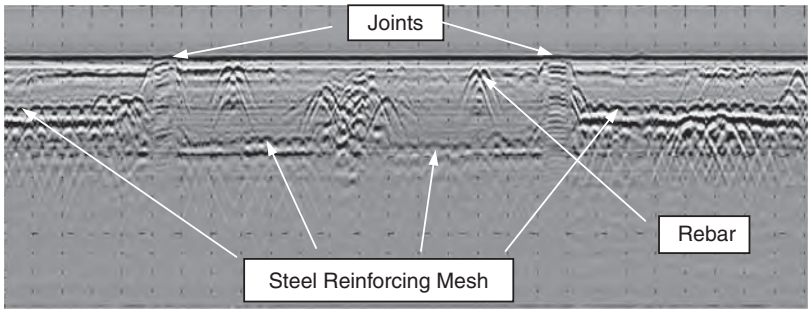


FIGURE 21.29 Typical radar image of concrete floor showing rears, joints, mesh, etc: 1 ns pulsewidth, horizontal markers every 10 cm, vertical scale 0.5 m (*Courtesy IEE*)

FIGURE 21.30 Temple steps (*Courtesy IEE*)

Archaeological applications of GPR have been varied, ranging from the exploration of Egyptian and North American Indian sites as well as castles and monasteries in Europe. The quality of the radar image can be exceptionally good, although correct understanding normally requires joint interpretation by the archaeologists and radar specialists. Since 1990, the Square Geophysical Survey Project, under the auspices of the National Museum of Scotland and the Glasgow Museums, have been carrying out geophysical and archaeological surveys at Square in Egypt. Square forms part of the necropolis of the ancient Egyptian capital city of Memphis. The burial grounds extend from Abu Roash, just to the north of Cairo, southward through Giza, Abusir, Saqqara, and Dahshur to Meidum approximately 20 km to the south. The famous Step Pyramid of the 3rd Dynasty ruler, King Zoser, dominates the site of Saqqara. The main monument is known as the Gisr el-Mudir, which consists of a 400 meter east to west by 600 meter north to south stone enclosure. The walls are of extremely crude construction, but massive. This monument may constitute one of the oldest stone buildings in Egypt and hence the world. One of the main goals of the project has been to determine what, if anything, lies within the enclosure. Despite many years of surveying, nothing has been found apart from a small area of mud brick pavement. A number of radar sections had been measured over this area previously, and in view of the magnetometry results, the radar profiles were reexamined. The sections have been measured on 25 meter centers, so it was purely by chance that one radar survey line went straight down the flight of steps (excavated) seen in Figure 21.30 with the radar image in Figure 21.31.

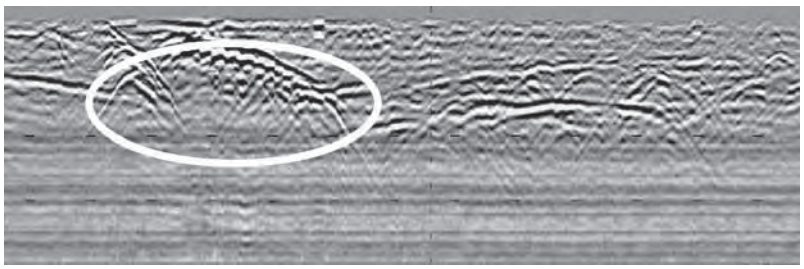


FIGURE 21.31 Radar section along the flight of steps (*Courtesy IEE*)

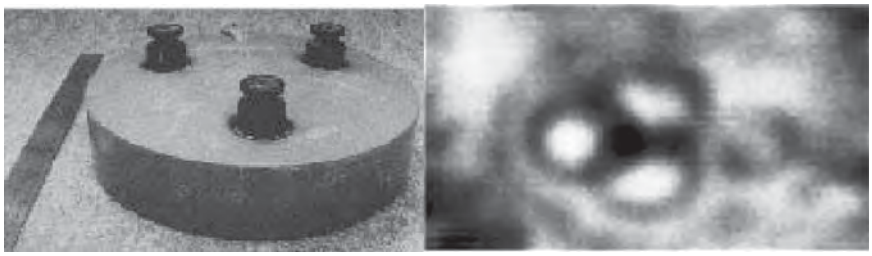


FIGURE 21.32 Example of GPR image of TMA3 AT mine (*Courtesy IEE*)

Abandoned antipersonnel land mines and unexploded ordnance are a major hindrance to the recovery of many countries from war. Their effect on the civilian population is disastrous and major efforts are being made by the international community to clear the problem. Most detection is done with metal detectors, which respond to the large amount of metallic debris in abandoned battlefield areas and hence have difficulty in detecting the minimum metal or plastic mine. GPR technology is being applied to this problem as a means of reducing the false alarm rate and providing improved detection of low-metal-content mines. Typical examples of radar images at various scales are shown in Figure 21.32 and Figure 21.33.

GPR has been used for surveying many different types of geological strata ranging from exploration of the Arctic and Antarctic icecaps and the permafrost regions of North America, to mapping of granite, limestone, marble, and other hard rocks as well as geophysical strata. The radar data shown in Figure 21.34 were collected at Finsterwalderbreen glacier, Svalbard, Norway, which is an island almost 80° north of Norway. The glacier is an 11 km long, land-terminated glacier with an area of 35 square km. The glacier depth starts at 30 meters and increases down to 250 meters. In the beginning of the profile, only the bottom echo is seen. At around 2 km horizontal distance, some internal scattering from the glacier is seen. This is scattering from the free water inside the glacier. From 4 km to 5 km, the bottom echo is difficult to see due to scattering.

The thickness of the various layers of a road can be measured using GPR techniques as shown in Figure 21.25. The great advantage is that this method is nondestructive and high speed (>40 km/hr) and can be applied dynamically to achieve a continuous profile or rolling map. The accuracy of calibration tends to reduce as a function of

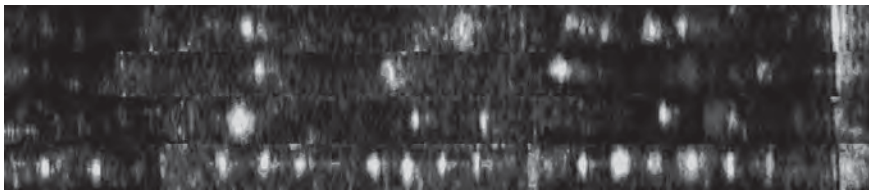


FIGURE 21.33 Example of AT mine images taken over a 4 m by 20 m test site with the MINDER GPR radar system (*Courtesy IEE*)

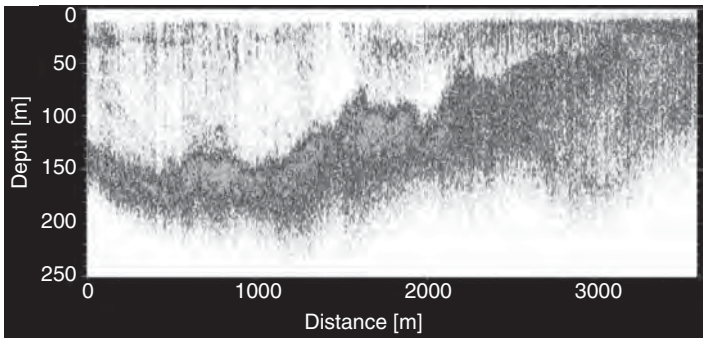


FIGURE 21.34 Radar profiles along the centerline on Finsterwalderbreen glacier at 320–370 MHz (*from Hamran et al., 2000*)

depth because of the attenuation characteristics of the ground. The accuracy may be quite high (i.e., a few millimeters) for the surface-wearing course, but will reduce to centimeters at depths of one meter.

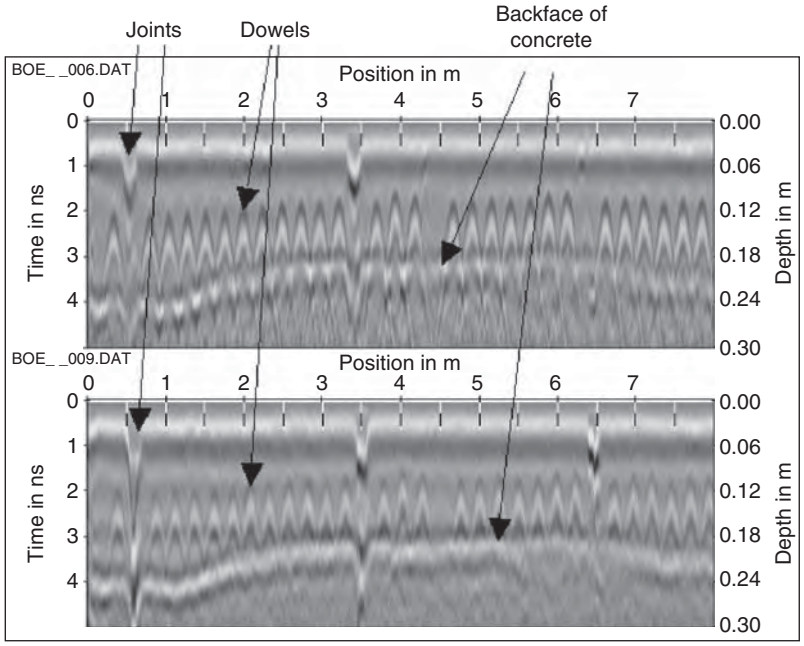


FIGURE 21.35 Radar images using 1.5 GHz pulse duration along an 8-m long transversal trace close to a joint of a highway concrete deck. Top: Polarization parallel to the dowels; bottom: Polarization perpendicular to the dowels. (*Courtesy IEE)*

FIGURE 21.36 UWB SAR image of buried AT mines in the Yuma Desert, aircraft at 400 m altitude (*Courtesy SRI International USA, Dr. R. Vickers*)

It has been shown that it is feasible to detect AT mines in dry soil conditions using airborne radar systems that operate over the frequency range 0.2–1 GHz. Figure 21.36 shows a radar image taken from an altitude of 400 meters above the Yuma desert. The radar operated at a depression angle of 45°and achieved a nominal resolution of 80 cm. It was capable of detecting metal AT mines of 30-cm diameter buried at a depth of 15–30 cm in a soil of conductivity 8–10 mmhos/m. Further details can be obtained from SRI International USA.

21.10 LICENSING

All countries require that GPR systems are properly regulated and operated in accordance with national and international requirements. Users should consult with their national authorities to determine the regulatory environment.

Within the European Union (EU), there are two main considerations governing the use of GPR. These are the use first of the equipment as a deliberate radio frequency radiator and second as an equipment that must satisfy the EMC requirements of the EU. The European Telecommunications Standards Institute (ETSI) regulatory body is in the process of drafting specifications and information can be found at http://www.etsi.org that will cover the use of such equipment as a deliberate radio frequency radiator. Legislation and an ETSI product specification means that this equipment will need to conform to the Radio & Telecommunications Terminal Equipment (R&TTE) — Directive. In the short term, until a new product specification is introduced and formally published in the Official Journal of the European Communities, the EMC Directive should be applied. All equipment, including ultrawideband radar or GPR, must be (**C**onformité **E**uropéene) CE marked to demonstrate that it satisfies the relevant directives of the European Union. The CE mark may only be applied when the requirements of all other relevant EU Directives, such as safety, have also been demonstrated. In the U.S. the FCC web site[30] provides current information and the limits are shown in Figure 21.37.

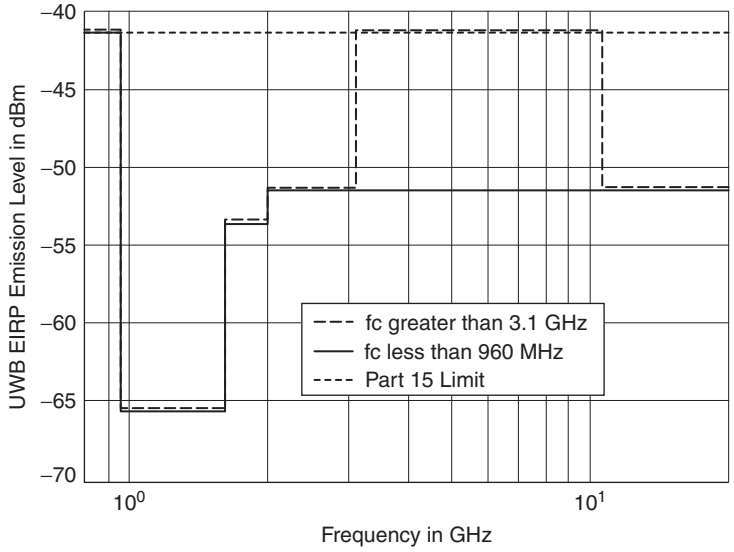


FIGURE 21.37 FCC emission limits

REFERENCES

1. D. J. Daniels, *Ground Penetrating Radar*, 2nd Ed. IEE Radar Sonar Navigation and Avionics Series, London: IEE Books, July 2004.
2. B. O. Steenson, "Radar methods for the exploration of glaciers," Ph.D. dissertation, Calif. Inst. Tech., Pasadena, CA, 1951.
3. S. Evans, "Radio techniques for the measurement of ice thickness," *Polar Record*, vol. 11, pp. 406–410, 1963.
4. R. R. Unterberger, "Radar and sonar probing of salt," in *5th Int. Symp. on Salt*, Hamburg (Northern Ohio Geological Society), 1979, pp. 423–437.
5. R. M. Morey, "Continuous sub-surface profiling by impulse radar," in *Proc. Conf. Subsurface Exploration for Underground Excavation and Heavy Construction*, Amer. Soc. Civ. Eng., 1974, pp. 213–232.
6. P. K. Kadaba, "Penetration of 0.1 GHz to 1.5 GHz electromagnetic waves into the earth surface for remote sensing applications," in *Proc. IEEE S.E. Region 3 Conf*, 1976, pp. 48–50.
7. J. C. Cook, "Status of ground-probing radar and some recent experience," in *Proc. Conf. Subsurface Exploration for Underground Excavation and Heavy Construction*, Amer. Soc. Civ. Eng., 1974, pp. 175–194.
8. J. C. Cook, "Radar transparencies of mine and tunnel rocks," *Geophys*., vol. 40, pp. 865–885, 1975.
9. K. C. Roe and D. A. Ellerbruch, "Development and testing of a microwave system to measure coal layer thickness up to 25 cm," Nat. Bur. Stds., Report No.SR-723-8-79 (Boulder, CO), 1979.
10. B. Nilsson, "Two topics in electromagnetic radiation field prospecting," Ph.D. dissertation, University of Lulea, Sweden, 1978.
11. D. J. Daniels, *Surface Penetrating Radar*, IEE Radar Sonar Navigation and Avionics Series 6, London: IEE Books, 1996.
12. See Reference 1.

13. Cook and Bernfeld, *Radar Signals, An Introduction to Theory and Application*, Norwood, MA: Artech House, p. 9.
14. M. Skolnik, *Radar Handbook*, 2nd Ed., New York: McGraw-Hill, 1990, Chap. 10.
15. F. Nathanson, *Radar Design Principles*, New York: McGraw-Hill, 1969, Chap. 8.
16. Wehner, *High Resolution Radar*, Chap 4.
17. Galati, *Advanced Radar Techniques and Systems*, IEE Radar Sonar Navigation and Avionics Series Vol. 4 London: IEE Books, 1993, p. 104.
18. Astanin and Kostylev, *Ultra-wideband Radar Measurements Systems*, IEE Radar Sonar Navigation and Avionics Series, Vol. 7, London: IEE Books, 1997, Chap. 1.
19. M. T. Tuley, J. M Ralston, F. S. Rotondo, A. M. Andrews, and AE. M. Rosen, "Evaluation of EarthRadar unexploded ordnance testing at Fort A.P. Hill, Virginia," *IEEE Aerospace and Electronics Systems Magazine*, vol. 17, issue 5, pp. 10–12, May 2002.
20. P. Debye, *Polar Molecules*, New York: Chemical Catalog Co., 1929.
21. K. S. Cole and R. S. Cole, "Dispersion and absorption in dielectrics, I, alternating current characteristics," *J. Phys. Chem*., vol. 9, pp. 341–351, 1941.
22. D. J. Daniels, "Resolution of UWB signals," *IEE Proc. Radar Sonar and Navigation*, vol. 146, pp. 189–194, August 1999.
23. C. Martel, M. Philippakis, and D. J. Daniels, "Time domain design of a TEM horn antenna for GPR," presented at Millennium Conference on Antennas and Propagation, April 2000.
24. D. J. Daniels, "An assessment of the fundamental performance of GPR against buried landmines," presented at SPIE Detection and Remediation Technologies for Mines and Minelike Targets XII, Orlando, FL, April 2007
25. G. Denniss "Solid-state linear FMCW systems –their promise and their problems," in *Proc. IEEE Int. Mic. Conf*., Atlanta, GA, 1974, pp, 340–345.
26. R. M. Narayanan, Y. Xu, P. D. Hoffmeyer, and J. O. Curtis, "Design, performance, and applications of a coherent random noise radar," *Optical Engineering*, 37(6), pp. 1855–1869, June 1998.
27. J. Sachs, P. Peyerl, F. Tkac, and M. Kmec, "Digital ultra-wideband-sensor electronics integrated in SiGe-technology," in *Proc. of the EuMC*, vol. II, Milan, Italy, September 2002, pp. 539–542.
28. J. Sachs and P. Peyerl, "Chip integrated UWB radar electronics," presented at Third DTIF Workshop, *Ground Penetrating Radar in Support of Humanitarian Demining*, JRC, Ispra, Italy, September 2002.
29. J. R. Wait, "Propagation of electromagnetic pulses in a homogeneous conducting earth," *Appl. Sci. Res. B*, vol. 8, pp. 213–253, 1960.
30. R. W. P. King and T. T. Nu, "The propagation of a radar pulse in sea water," *J. Appl. Phys*., 73, (4), pp. 1581–1589, 1993.
31. D. B. Rutledge and M. S. Muha, "Imaging antenna arrays," *IEEE Trans*, AP130, (Q), pp. 533–540, 1982.
32. Federal Communications Commission, http://www.fcc.gov/aboutus.html.

Chapter 22
Civil Marine Radar

Andy Norris

22.1 INTRODUCTION

In terms of the number of systems in worldwide use, civil marine radar (CMR) is the largest radar market of all time. The number of vessels of all types currently fitted with radar probably amounts to around 3 million, but there are no official records to verify this estimate.

CMR breaks down into two main application areas. The vast majority are used at sea and on navigable waterways by ships and smaller craft; the others are used by port and coastal authorities for vessel surveillance from land-based sites. The latter group are normally known as *vessel tracking service (VTS*†*)* radars. Radars are available for leisure craft, fishing vessels and merchant ships, and all operate either in the 3 GHz or 9 GHz bands. Many navies also use standard or specially modified CMR for navigational purposes. Not only does it provide a suitable navigational tool but also its transmissions are identical to conventional commercial traffic, allowing safe navigation without necessarily highlighting a vessel's military purpose.

The biggest influence on the requirements of shipborne CMR comes from the International Maritime Organization[1] (IMO). A United Nations agency based in London, IMO is concerned with international maritime safety and the protection of the marine environment. In particular, IMO issues requirements and guidelines on the installation and use of radar equipment on commercial ships. These are vigorously enforced by the laws of individual maritime states. The purpose of shipborne radar, as defined by IMO, is to "assist in safe navigation and in avoiding collision by providing an indication, in relation to own ship, of the position of other surface craft, obstructions and hazards, navigation objects and shorelines."[2] The International Association of Marine Aids to Navigation and Lighthouse Authorities[3] (IALA) recommends operational and technical requirements for VTS radars.

This chapter explains the special requirements of CMR, both from a practical and a regulatory point of view, and looks at the technology and system concepts that are being used to meet these requirements. Until the first decade of the present century, CMR shipborne technology had been solely based on magnetrons as the basic source of transmitted power. Since 2004, IMO has encouraged the use of coherent radar solutions in an attempt to improve the detection of targets in heavy sea clutter conditions.

† A list of all used maritime abbreviations is included at the end of the chapter.

In the marine world, these have been called *New Technology Radars*. They are permitted to transmit any waveform at 3 GHz, providing the spectrum limitations on marine radar are not exceeded.[4] The limits have been agreed within the International Telecommunication Union[5] (ITU), a United Nations agency based in Geneva.

This chapter concentrates on the requirements and design of radars for commercial ships normally in excess of 300 gt (gross metric tonnage), where radar fitment is compulsory and highly regulated. Worldwide, there are about 50,000 of these vessels, and many are required to carry more than one radar. Three or even more radars are sometimes carried voluntarily by large ships. Radar forms an important part of a vessel's total navigation equipment fit. Increasingly, the bridge of a ship is designed as an integrated concept, covering navigation, communications, engine control, and cargo monitoring facilities. Figure 22.1 illustrates a modern *integrated bridge system* (IBS) as fitted on a cruise ship. The radar displays are seen to form a prominent part of the system. Radars fitted on smaller fishing vessels and leisure craft share many of the features of radars designed for ships but are necessarily more compact; a typical small boat radar is shown in Figure 22.2. Specific requirements for these radars, where they differ to any extent from the design of shipborne radars, are discussed within relevant parts of the chapter. Radar for VTS is separately covered in Section 22.10.

The challenges facing designers of shipborne radar are detailed within Section 22.2. These radars have to meet certain international standards, which are discussed in Section 22.3. Section 22.4 concentrates on the technology, and Section 22.5 looks at target tracking. Radar targets are being increasingly displayed with electronic chart data as an underlay. This is outlined in Section 22.6, together with other user interface issues. Section 22.7 looks at the links between radar and the relatively new *Automatic Identification System* (AIS), which replicates some functions previously provided solely by radar. Marine radar beacons, including racons, *Search and Rescue Transponders* (SARTs), and *Radar Target Enhancers* (RTEs), are described in Section 22.8. There is a short discussion in Section 22.9 on shipborne radar performance validation testing.

FIGURE 22.1 Ship's integrated bridge system (*Courtesy of SAM-Electronics GmbH*)

FIGURE 22.2 Small boat radar (*Courtesy Furuno USA, Inc.*)

Shipborne radar has had a remarkably long history. Its conception in the period from 1945 to 1948 was remarkably prophetic and still reverberates into the present century. For this reason, a short Appendix to this chapter outlines the early steps in the evolution of global standards for shipborne radar.

22.2 THE CHALLENGES

Environmental. Civil marine radar, particularly shipborne navigation radar, is a surprisingly demanding application. The radar head on a CMR comprises the antenna and turning gear, the receiver down to IF or to digital format and often the transmitter. It has to operate in extreme environmental conditions over an extended temperature range (down to –40°C in some parts of the world); in high levels of wind, vibration, and shock; and also in heavy precipitation and salt water spray. Even within the normally benign conditions on the bridge of a large modern ship, the display and radar processor can be subject to high levels of shock and vibration and must meet high variations in temperature (–15°C to +55°C). On small craft, the display and radar processor are often fitted in minimally enclosed areas and are subject to very damp and salty conditions. In these environments, the radar has to detect targets that can have echoing areas ranging from less than one square meter to many tens of thousands of square meters; important targets can have relative speeds ranging from stationary to 100 knots or more; the targets can be situated in extreme precipitation and sea clutter conditions; and the radar antenna is not mounted on a static nor a stable platform. The radar is used to prevent collisions and groundings at sea and is, therefore, an important safety related system, requiring integrity and reliability. For most commercial ships, the radar needs to meet stringent, internationally agreed performance criteria. Despite these requirements, systems are sold in a highly competitive market and prices are, therefore, kept keen. Prices range from around $10,000 for a complete but basic 9 GHz system to $40,000 and above for a fully featured 3 GHz system. Radar systems suitable for the leisure market can sell for less than $1,500.

TABLE 22.1 IMO Required Detection Performance in Clear Conditions (*Courtesy of IMO*)

	Target Feature	Detection Range in NM (for specified target size)		
Target Description	Height Above Sea Level (meters)	9 GHz NM	3 GHz NM	Target Type
Shorelines	Rising to 60	20	20	Distributed
Shorelines	Rising to 6	8	8	
Shorelines	Rising to 3	6	6	
SOLAS ships† (> 5,000 gt)	10	11	11	Complex
SOLAS ships† (> 500 gt)	5.0	8	8	
Small vessel with radar reflector meeting IMO performance standards	4.0	5.0 (7.5 m^2)	3.7 (0.5 m^2)	Point
Small vessel of length 10 m with no radar reflector	2.0	3.4 (2.5 m^2)	3.0 (1.4 m^2)	Complex
Typical navigation buoy	3.5	4.6 (5.0 m^2)	3.0 (0.5 m^2)	Not specified, point target assumed
Navigation buoy with corner reflector	3.5	4.9 (10 m^2)	3. 6 (1.0 m^2)	Point
Typical channel marker	1.0	2.0 (1.0 m^2)	1.0 (0.1 m^2)	Not specified, point target assumed

† Ships conforming to the IMO Safety of Life at Sea (SOLAS[9]) regulations

Detection Performance. In-the-clear detection requirements for such radars are not particularly demanding. An 80% probability of detection and a probability of false alarm of 10^{-4} is specified by IMO, as shown in Table 22.1.[2]

Taking into account all performance requirements, typical compliant systems for commercial vessels have peak transmit powers of 4 – 60 kW, the lower powers being confined to 9 GHz systems. Antenna gains from 28 to 33 dB are typical, with associated horizontal beamwidths ranging from about 2.5° to less than 1°. Pulse lengths are switchable, generally in the range from 50 ns to 1 μs, with PRFs ranging from 350 to 3,000 Hz or more. Pleasure craft systems typically have peak powers of 2 – 4 kW and utilize antennas with horizontal apertures as small as 450 mm and gains of about 24 dB. These all operate at 9 GHz. The greatest technical challenge in designing marine radars is to maintain good target detection in high levels of sea and precipitation clutter.

Precipitation Clutter. It is well known that circular polarization (CP) can be a counter to rain clutter because its reflection is predominately cross-polarized to incident circular polarization. However, very few shipborne navigational radars use CP, even though 10–20 dB improvement in rain clutter rejection is typically achieved from its application. Two important factors have contributed to this. First, it makes the antenna more expensive, exacerbated by IMO's requirements that 9 GHz radars must be at least switchable to horizontal polarization when searching for survival craft fitted with Search and Rescue Transponders (see Section 22.8). Second, by use of small range cells and by implementing conventional signal differentiation techniques, modern radars, particularly at 3 GHz, give reasonable performance in most commonly experienced precipitation clutter. Therefore, users and maritime authorities are generally satisfied with the performance of linearly polarized systems in precipitation.

Because precipitation clutter is distributed in a relatively uniform manner, passing the received waveform through a differentiator gives prominence to targets embedded within the clutter by ensuring that the average clutter level is kept well below saturation. The differentiator has minimal effect on normal targets because of their small extent in the time domain. This means that target visibility is improved. It should be noted that this technique does not give subclutter visibility. This process is conventionally named *Fast Time Constant* (FTC). The time constant of the differentiator is operator-adjustable with a so-called rain clutter control, which allows the target-to-clutter ratio to be optimized for the particular precipitation scenario.

The vertical pattern of a shipborne navigational radar antenna needs to be relatively wide to cope with the ship's pitch and roll, which is assumed to be a maximum of ±10°. (Use of a stabilized platform would not meet the market's price demands.) This limits the vertical beam-shaping that can be used to reduce both precipitation clutter and vertical lobing effects. However, the relatively short range of most targets of real interest means that the volume of the radar-illuminated precipitation is relatively low, helping to make the clutter rejection of the differentiator adequate for its purpose. Such clutter varies approximately with the fourth power of frequency, and so a 3 GHz system inherently experiences 19 dB less clutter than a 9 GHz system, assuming identically sized clutter cells. For this reason, on ships fitted with both 3 and 9 GHz radars, the 3 GHz radar is often preferable, except when maneuvering in close situations, for example, in harbors when the normally superior azimuth resolution of a 9 GHz radar is preferred.

Sea Clutter. The reduction of sea clutter to levels acceptable to the user is a far more difficult problem, and as yet, commercial radars do not meet all the ideal demands of users. Small craft and buoys can easily be obscured in sea clutter. In the days before precise *Global Navigation Satellite Systems* (GNSS), such as GPS, the safe navigation of a ship in coastal waters in poor visibility was dominated by the radar being able to discern navigation markers, such as buoys. Passive markers, including those supplemented by radar reflectors, can be notoriously difficult to detect in higher sea-states, and therefore, some markers are supplemented by radar beacons (called racons—see Section 22.8). Racons are relatively expensive and need maintenance in often difficult to access locations, so their use is restricted. While GNSS, enhanced by the growing use of electronic charts, has helped greatly in informing mariners of the precise position of their vessel, radar is still used as an important secondary source of position. Reliance on GNSS alone has been at the root of many marine accidents.

The main use of marine radar is to assist collision avoidance. Visual observation and radar remain the primary methods for determining the risk of collision with other vessels and also with floating debris and ice. Automatic Identification Systems (AIS—see Section 22.7) offer potential to assist with collision avoidance of cooperative targets, but it cannot be assumed that all vessels are fitted with AIS, particularly small craft, or that a target vessel's AIS is operational.

The traditional way for mariners to optimize their radar for detection of targets in sea clutter is by careful adjustment of the "gain" and "sea-clutter" controls. The gain control effectively alters the detection threshold. On a modern marine radar, the sea-clutter control is best described as a method for adjusting the shape of the radar's *sensitivity time control* (STC) in order to match it with the present level of clutter returns. STC is often also called *swept gain*. The STC law, and the way it varies by use of the manual control, can be complex. It is attempting to reduce the dynamic range of the received waveform and to provide, in association with the gain settings, optimized thresholds.

Nowadays, it normally involves sophisticated adaptive thresholding techniques, which are discussed in Section 22.4.

Although this helps to set the threshold to appropriate levels, it does not remove the intrusive "spiky" component of sea clutter that can make wanted targets difficult to observe. However, over a typical antenna scan time of a marine radar (2–3 seconds), the spikes are normally decorrelated, whereas returns from targets are generally correlated—therefore, the application of scan-to-scan correlation can improve the target-to-clutter ratio, but it will also remove weak and fast-moving targets. Many years ago, Croney[6] showed that significant improvements in detecting small targets in sea clutter could be obtained by ensuring that integration was performed at intervals longer than the decorrelation time period of the sea clutter. He used an antenna rotating at up to 600 rpm and a PRF of 5 KHz. This gave two correlated pulses per beamwidth, but the pulses from the next scan, 0.1 s later, were decorrelated from the former.

Crony noted that the rapid scanning of the antenna allowed the eye/brain functions of the operator to perform scan-to-scan correlation. Although it is easy for modern systems to perform this correlation digitally, the difficulties in having an antenna rotating at this speed (mainly now a cost issue) have prevented this from becoming established practice. However, more recent work in Canada has resurrected this idea for detecting floating ice hazards, where antenna rotation rates of 500 rpm and PRFs of 12 KHz have been proposed.[7] Terma A/S, a Danish company involved in supplying high performance marine radars, mainly in the noncommercial market, produces the Scanter™ radar, which has an option that simultaneously transmits on two frequencies from a squinting slotted waveguide array. This produces two beams separated by a few degrees in azimuth. The temporal beam separation is such that sea clutter can be decorrelated between the beams, further enhancing the detection of targets in clutter. This technique could potentially be used by solid-state CMRs. (Section 22.4).

On a conventional shipborne radar, an experienced operator can manually set the detection threshold to give the best setting over any given area, but this is often only effective over a small proportion of the total radar image. The use of automatic thresholding is able to give better detection over a complete scan but often cannot compete with a skilled operator optimizing detection over a restricted area. In some conditions, no existing radar gives the performance that a user ideally needs, despite the use of 50 ns pulses and sophisticated clutter processing technology.

Sub-clutter visibility is potentially obtainable from coherent CMRs, which have been made affordable by the continued reduction in cost of microwave power semiconductors (for instance, using Gallium Nitride technology), precision digitally controlled signal generators, and fast digital signal processors. Coherent CMRs are discussed under "Solid-state CMR" in Section 22.4.

Vertical Lobing. Clutter is not the only cause of detrimental performance of marine radar. Direct reflections from a target arrive at the radar antenna and combine vectorially with target reflections that have also been reflected by the sea's surface. This effect produces a summed signal at the radar antenna that is a function of both the target height and the radar antenna height above the sea, as these affect the path length difference of the direct and reflected radiation. Obviously, the effect is reciprocal for both transmit and receive paths. For a point target and a sea of defined roughness, the calculation to determine the resultant effect is relatively straightforward and results in the classic lobing pattern (see, for example, Briggs[8]). For a target with reasonable vertical extent, such as a ship, the lobing structure becomes very complex and is less likely to produce troublesome nulls. However, the detection of a small target, such as

a buoy or a pleasure craft, enhanced by a radar reflector can create significant vertical lobing effects that can be a problem to the user. In particular, in very calm seas pronounced nulls can be experienced, and for users, it can be disconcerting when a target clearly evident from the bridge window is not visible on the radar display, despite the apparently good conditions.

Since smooth seas can also be associated with mist and fog, vertical lobing effects can become a significant problem because the reduced visibility means that radar often becomes the sole method of detecting other vessels. The absence of sea clutter gives the user a false sense of security that all targets will be easily visible. On ships fitted with both 3 and 9 GHz radars, frequency diversity becomes very useful as the spatial frequencies of the vertical lobes for the radar are different. Surprisingly, even though some radar companies provide the option, very few ships have facilities that allow the 3- and 9 GHz signals to be combined into a single radar display in an automatic process, maximizing the benefits of frequency diversity. Some large ships have an additional 9 GHz radar mounted on the bow at deck level. This has two advantages. First, the vertical lobing will have a different (lower) angular frequency to the main, high-mounted 9 GHz radar. Second, its performance in sea clutter will be enhanced, as the grazing angle to the sea will be closer to horizontal, thereby lessening the reflection coefficient of the clutter. The long range performance of the auxiliary system is, of course, compromised by its low position.

Moving Platform. A particular complication of shipborne radar arises because the antenna is mounted on an unstable moving platform. This movement has six components—three translational and three rotational, all of which are typically varying. The motions are complex; the translational components are surge, sway, and heave, and the rotational are roll, pitch, and yaw. Components can be quasi-harmonic when caused by wave motion. In practice, the ship is navigated on notions based on course, heading, and speed in Earth and sea-fixed coordinate systems. The additional wave-induced motions can produce uncompensated errors in radar-derived information, which add to any measurement errors in the course, heading, and speed of the vessel. This affects the precision obtained when displaying radar-derived data, which can differ when switching between the various radar stabilization modes used on CMR. For instance, in order to facilitate both collision avoidance and position fixing activities, shipborne radar displays have always had two particular stabilization modes: Head-up and North-up. The "up" direction refers to the vertical (y-axis) direction of the radar display; "Head" refers to the ship's heading. Head-up maximizes the relationship to the visual scene, and North-up aids comparison with paper charts. Nowadays, Course-up is also provided, as this eliminates small oscillations in the radar image due to the ship's yaw that occur when the display is set to Head-up. Each of these directional modes can be set with target tracking vectors shown as relative to the motion of the ship, the ground, or the average sea motion.

22.3 INTERNATIONAL STANDARDS

Spectrum use aspects of all radars, including frequency band of use and RF emission constraints, are controlled by the ITU.[5] Following ITU requirements, marine radars are permitted to operate in the 9.3 to 9.5 GHz band (X band) and in the 2.9 to 3.1 GHz band (S band).

The IMO International Convention for the Safety of Life at Sea (SOLAS)[9] is an established and accepted set of principles and rules aimed at ensuring that ships meet certain requirements to enhance both safety and protection of the environment. The member governments (flag States) of IMO have agreed that SOLAS requirements are embodied within their national maritime laws and regulations. Within Chapter V of *SOLAS - Safety of Navigation*, the requirements for the carriage of navigation equipment are defined. These vary according to the size and purpose of a ship. All passenger ships and all ships above 300 gt need to carry at least one radar with tracking facilities.

Footnotes within Chapter V of SOLAS identify the recommended IMO performance standards with which the equipment should conform. IMO has had recommended radar performance standards[2] since 1971, published as annexes to IMO Resolutions. However, by 1980, radar manufacturers were reporting difficulties because differing interpretations by national maritime administrations meant that radars had to be specifically designed to meet individual flag State requirements. The level of technical detail required to remedy this was outside the remit of IMO, and it was agreed that a Technical Committee (TC80) within the International Electrotechnical Commission[10] (IEC) would determine technically based interpretations of IMO radar performance standards. In addition, it was agreed that the IEC standards would include test procedures, which could be used by national maritime administrations (such as the Coast Guard in the United States of America) to test for conformance of specific designs by manufacturers to IMO and ITU requirements. Today, virtually all national administrations use IEC standards to assess radar and most other IMO-defined navigational and radiocommunications equipment.

IMO performance standards and the SOLAS Convention are regularly revised, so it is important to check the current status of the standards. IEC 62388[11] defines the technical and test standards based on the IMO radar performance standards. IEC standards also undergo regular revision. An average radar installation has a life normally exceeding 10 years, so radars designed and approved to previous standards will continue to be used for some years after new standards have been put in place. Retrofitted equipment must meet the latest standards.

IMO radar performance standards previous to those in force on 1 July 2008 require compatibility with existing racons (radar beacons) and at 9 GHz, Search and Rescue Transponders. This implies the continued use of short pulse radars. However, for the 2008 standards, IMO has encouraged improvements in sea clutter performance by dropping the need for racon compatibility at 3 GHz, thereby allowing other forms of modulation that would enable affordable coherent processing techniques. Because all ships above 300 gt need to carry at least a single 9 GHz radar, it means that racon (and SART) detection capability is maintained. This approach gives IMO an indefinite period to assess the impact of the new regulations on the detection of targets in sea clutter before deciding what should happen with radar, racons, and SARTs at 9 GHz.

Another major change in the requirements of older standards is that all new radars must include provision to display Automatic Identification System (AIS) targets and that their related information can be accessed on the radar display. The requirements for target tracking has also had a major revision, with automatic tracking facilities being required for all radars. The integration of electronic chart data as a background to radar images is also embodied within IMO standards. Radars with this optional facility are known as *Chart Radars*.

The minimum detection performance required in clear conditions is tabulated in Table 22.1. Measurements of range have to be within 30 meters accuracy (or within 1% of the maximum range scale in use) and within 1° bearing (azimuth angle). Navigational buoys with the characteristics given in Table 22.1 have to be detectable at a minimum range of 40 meters. Two "point" targets on the same bearing have to appear as two distinct targets if they are separated by more than 40 meters in range. A 2.5° azimuth resolution is also required. All these performance figures are considered to be peak errors, which can be assumed to mean 95% values, measured with standardized point targets. The IMO performance standards recognize that the detection performance of radars working within conditions of clutter will not necessarily give the performance defined for clear conditions. Manufacturers are required to provide effective manual and automatic anti-clutter functions and must specify the expected degradation in rain at 4 mm and 16 mm per hour and for sea states 2 and 5, including combinations of sea and rain clutter.

Radars designed for conventional vessels need to operate with relative speeds up to 100 kt. For high-speed craft, such as multi-hull fast ferries, the radars need to operate with relative target speeds up to 140 kt. Older standards required a minimum antenna rotation rate of 20 rpm, but this explicit requirement has been omitted from the new standards as other dependent requirements are adequately specified, such as the maximum relative speed of targets and tracking accuracies. The IMO performance standards specify that radar equipment should meet the environmental requirements and test procedures defined within IEC 60945.[12] This is a comprehensive set of requirements that are applicable to all ships' navigational and radio communications equipment. They cover such aspects as temperature, shock, vibration, corrosion, and resistance to water and oil ingress. Detailed requirements on electromagnetic emissions and immunity to the electromagnetic environment are mandated. IEC 60945 also specifies general requirements on ergonomics, software development, and safety. A further set of IEC standards contained within the IEC 61162 series,[13] define the messages used for navigation and radiocommunications equipment to interchange digital data. A shipborne radar is likely to be receiving messages from many items of navigation equipment, such as AIS, GPS, gyrocompass, log, and echo sounder, and is also likely to be communicating track information to electronic chart systems and possibly other radar displays.

A number of manufacturers produce radars specifically designed to be used on vessels using the world's major inland waterways. These are known as *river radars*. They are epitomized by their superior short-range performance and by having a display in "portrait" format in order to get maximum look-ahead along the waterway. The maximum display range on the shortest scale is typically 150 meters. These radars are normally designed to meet requirements for radars on vessels navigating the River Rhine.[14]

Radars for the fishing and leisure markets are not covered by SOLAS. Until 2004, such radars did not have an internationally recognized standard for manufacturers to follow. IEC 62252[15] is now the agreed international radar standard for "craft not in compliance with IMO SOLAS Chapter V" and was initially issued at the instigation of manufacturers. Now an increasing number of national maritime administrations are insisting that all new small craft radars sold in their jurisdiction conform to this standard. IEC 62252 recognizes three classes of radar. Class A is intended for commercial craft under 150 gt; Class B is for recreational craft; and Class C is for small recreational craft. The main performance requirements are detailed in Table 22.2.

TABLE 22.2 Radar Performance Requirements for Small Craft (*Courtesy of IEC*)*

Class	Beamwidth	Minimum Display Size	Coast-Line Detection Range: Rising to 60 m	Coast-Line Detection Range: Rising to 6 m	Point Target Detection Range: 400 m^2 7.5 m ht	Point Target Detection Range: 10 m^2 3.5 m ht	Point Target Detection Range: 5 m^2 3.5 m ht
A	≤ 4.0°	≥ 150 mm	9 nm	5 nm	5 nm	2 nm	1 nm
B	≤ 5.5°	≥ 85 mm	5 nm	3 nm	3 nm	1 nm	N/A
C	≤ 7.5°	≥ 75 mm	5 nm	3 nm	3 nm	1 nm	N/A

22.4 TECHNOLOGY

Antennas. Antenna maximum sidelobes are specified for SOLAS and non-SOLAS radars in IEC 62388[11] and IEC 62252,[15] respectively. These are summarized in Table 22.3.

On non-SOLAS radars, the antenna rotation rate is specified to be not less than 20 rpm but is not directly specified for SOLAS-approved radars. In practice, antennas for existing shipborne radars normally rotate at 25–30 rpm; on high-speed craft, the rotation rate is typically 40–45 rpm. For SOLAS vessels, the antenna must be able to start and operate in relative wind speeds up to 100 kt; other environmental requirements for the antenna system are detailed in IEC 60945,[12] where there are specific tests for "exposed" equipment.

There are no explicit requirements on other antenna parameters for SOLAS-approved systems, such as beamwidths and gain, but these obviously need to be compatible with the total radar performance requirement. For instance, azimuth resolution has to be better than 2.5°; target bearing has to be determined to within 1°; and the system must operate in conditions when the ship is rolling and pitching ±10°. Typical antenna gains and beamwidths have been outlined in Section 22.2.

From the 1960s onwards, the use of a slotted waveguide linear array, mounted in a linear flared horn has been the most common antenna solution for shipborne radars. Because at least horizontal polarization has to be provided on 9 GHz SOLAS radars, slotted array solutions generally have their slots cut into the narrow wall of the horizontally mounted waveguide. A vertical slot (perpendicular to the waveguide edge) couples no power, but as the slot is increasingly angled, more power is coupled out. The slots are normally of a resonant length (half-wavelength) to couple out sufficient power. This extends the slot into the broad wall of the waveguide, but it also makes them easier to construct—conceptually by a sawing action into the narrow wall. Residual power at the end of the array (typically less than 5%) is dissipated into a matched load. Conventionally, the array is end-fed, although center-fed alternatives are sometimes used.

TABLE 22.3 Antenna Sidelobe Performance Requirements (*Courtesy of IEC*)

Radar Class	Maximum Sidelobe Level (dB): Within ±10°	Maximum Sidelobe Level (dB): Within ±10°
SOLAS	–23	–30
Non-SOLAS Class A	–20	–23
Non-SOLAS Classes B & C	–18	–19

* IEC 62252 ed.1.0 Copyright © 2004 IEC, Geneva, Switzerland. www.iec.ch.

If slots were spaced at the guide wavelength in an attempt to get an equi-phase wavefront at the antenna face, then large grating lobes would be generated in the far-field pattern. This would occur because they would be spaced at a free-space distance of more than one wavelength. To overcome this, slots are spaced at nominally half the guide wavelength but are angled alternately to the vertical in order to induce the necessary phase reversals. In practice, the slots are placed slightly away from half the guide wavelength spacing to avoid slot-generated mismatches in the waveguide becoming resonant. This creates a tilted phase front across the array, which causes the beam to squint to an angle that is frequency dependent. Individual manufacturers produce radar systems operating over a restricted band, much less than the overall radar band, which removes any need for individual squint compensation when magnetrons are replaced. The required sidelobe performance is not demanding (Table 22.3), and so simple aperture distributions, such as pedestal-based cosine squared, are common. The small vertically polarized fields produced by each slanted slot need to be suppressed as they can otherwise lead to high cross-polarized sidelobes from the array, exacerbated by the phase reversal of the cross-polar component from slot to slot, causing cross-polar grating lobes. This can be achieved with a printed polarization filter in front of the array or by effectively creating, as part of the structure, a short length of open-ended waveguide in front of each slot, with dimensions that make it below cutoff for vertical polarization.

Slot characterization is normally performed by measurement, rather than by detailed electromagnetic analysis. This allows all construction details, including those required for polarization filtering, to be incorporated into the slot characterization; sufficient accuracy is difficult to achieve using numerical analysis. Provided the characterization is done carefully and good manufacturing techniques can guarantee tolerances, it is reasonably straightforward to produce affordable antennas that meet IMO requirements. The required vertical beamwidth is normally obtained by a linear flare. The flare angle is chosen such that it creates a reasonably phase-constant vertical distribution at its aperture. The vertical amplitude distribution approximates to a cosine because of the horizontally polarized field. Vertical beamwidths are typically about 25° wide at the 3 dB points.

Not unnaturally, cost, for a given performance, is the prime driver in the system designer's choice of antenna. While the conventional slotted linear array is widely used, there are examples where different cost tradeoffs have been made. For instance, the use of a dielectric block mounted directly in front of the slotted waveguide array, in place of the flared section, has been used as an alternative. The leaking energy from the top and bottom faces of the dielectric block adds with energy emerging from its front face, giving forward gain. It is the depth dimension of the dielectric block that determines the gain, somewhat analogous to the length of a Yagi antenna. This effect reduces the height of the antenna compared to a conventional design by about a factor of three, typically from about 300 mm to 100 mm at 3 GHz. This means that there is considerably less wind-loading. The dielectric constant of the block can be quite low, which, with the reduced wind-loading, results in a very lightweight structure. This saves costs in the antenna turning gear and makes the installation easier. An example of this type of antenna, produced by Kelvin Hughes, is illustrated in Figure 22.3.

Small craft radars have used printed arrays for some years, as well as slotted waveguide arrays. Small, horn-fed parabolic reflector systems have also been used. Antennas for small craft are normally housed within a radome, which protects the antenna and up-mast electronics of the radar environmentally and prevents the antenna from snagging the rigging. In particular, it assures that there is no danger to

FIGURE 22.3 Low-profile 3.9 meter S-band shipborne radar antenna (*Courtesy of Kelvin Hughes Ltd.*)

users from rotating mechanisms, since the radar head can be mounted in areas open to human access. Antennas using printed arrays use integral-printed power dividers. The arrays are usually two-dimensional, dispensing with the need for a flared section, and normally consist of radiating patches, rather than printed dipoles. Horizontal apertures of 450 and 600 mm are common. Printed technology is not generally used for shipborne antennas; slotted waveguide arrays remain the cost-effective solution for larger arrays, particularly as the higher power creates additional complications for printed power dividers.

Poor antenna siting is a common cause of radar performance degradation on ships as well as on smaller craft. Particularly on ships, it is surprising that installations are still being implemented that create significant blockage to the radar. Blind arcs are common from funnels and other superstructure, and there can be significant sidelobe degradation due to smaller structures, such as VHF antennas, causing blockage close to the radar antenna aperture.

RF Head. The RF head normally comprises the transmitter and the receiver down to IF or digital baseband, as well as the antenna and turning gear. Its design for both magnetron-based SOLAS and non-SOLAS radars follows conventional principles. The magnetron is connected to the antenna via a duplexer and a rotating joint. The magnetron has a typical operational life of about 10,000 hours and is by far the lowest lifed component in the whole system. The duplexer is nowadays a three- or four-port ferrite circulator. Use of a four-port device is preferred as it presents a better matched load to the magnetron and, therefore, gives a cleaner RF spectrum. The *low-noise front-end* (LNFE) subsystem is connected to the circulator via a PIN diode limiter, which protects the LNFE during pulse transmission.

The modulator to the magnetron is typically a *pulse forming network* (PFN), basically comprised of capacitors and inductors. Operator control of the pulse length effectively switches in different choices of reactive components. The discharge of the PFN is controlled by a high voltage switch, which is often a silicon-controlled rectifier; thyristors and FETs are also used. FET modulators are sometimes driven directly by a pulse input rather than a PFN. Finally, a pulse transformer matches the PFN to the impedance seen at the magnetron cathode. A pulse of around 10 kV is required to fire the magnetron. In order to get good performance over a wide range of pulse lengths, designs utilize a great deal of empirically derived knowledge, and actual circuits can be surprisingly complex. A 50 ns pulse hardly achieves any period of stability—rise times are normally restricted to about 10 ns in order to limit out-of-band interference,

and fall times are usually longer. The extra high voltages involved can lead to poor reliability if the design does not adequately address the associated problems. Careful physical layout is essential and consideration must be given to the effects of operating in a potentially damp environment. Pulse timing can be purposely jittered on a pulse-to-pulse basis. Pulse-to-pulse correlation in the radar processor then very effectively blocks interference from other radars, albeit with a small but generally acceptable degradation in detection performance.

After the limiter, the LNFE is preceded by a bandpass filter to reduce the effects of out-of-band interfering signals. The LNFE consists of an RF amplifier (typically giving about 10 dB gain) and a balanced mixer, local oscillator, and IF head amplifier. These are normally supplied to the radar manufacturer by specialist companies as a complete subunit. The overall system noise figure is typically 4 to 5 dB, but lower figures are achievable. The frequency of the local oscillator is generally driven by a control signal, derived from within the IF amplifier. It includes facilities for manual frequency control by the operator. The latter can be useful, for instance, when looking for SARTs in heavy sea clutter, as it allows returns from own transmissions to be desensitized. The local oscillator uses a Gunn diode or an FET, giving a typical intermediate frequency of 60 MHz.

Detection and Processing. After the LNFE, a logarithmic amplifier reduces the dynamic range of the received signal to prevent limiting. A dynamic range of around 100 dB is typically achieved in perhaps an eight-stage amplifier. Filtering, consistent with the transmitted pulse length, is applied within the IF amplifier. The output from the log amplifier enters a diode-based envelope detector, converting the signal to baseband for subsequent threshold processing. Setting the threshold has become intimately connected with the sensitivity time control (STC) of the radar. The fundamental use of STC is to take out the distance-related dynamic range of received signals. At close range, STC classically follows an inverse fourth power law, merging to an inverse cubic law in the region where sea clutter dominates, in accordance with basic theory. Because transition range is a function of antenna height, a setting for this may be needed when the system is originally installed. The operator's manual sea-clutter control is used to adjust the transition range. Nowadays, the form of the STC curve and the effects of the manual control are based on the practical experience of individual manufacturers, which contributes greatly to the actual effectiveness of a particular radar in sea clutter. Even under manual control, the detailed shape of the STC curve may have a complex adaptive element to it, in order to optimize thresholds over a broader range.

Under automatic settings, thresholding becomes increasingly sophisticated but often still allows some manual optimization. The curve may adapt to internal calculations made on returns from the last pulse or from a succession of pulses. It may also include more complex clutter-mapping processes. All of this is attempting to create a constant false alarm rate across the radar display. Since the radar is sited on a moving platform, subject to complex dynamics, difficulties arise with clutter mapping. However, modern processors permit affordable thresholding algorithms of considerable complexity. Manufacturers keep their own processes highly confidential because of the effort that has gone into their empirical optimization. Theoretical clutter models have been generally found unsuitable to be used for optimization. Even optimizing for sea conditions found in one particular area can create suboptimal solutions in other areas, and therefore, data from a number of regions need to be used to design a globally effective product.

STC and thresholding is a process that is controlled in the digital domain but often applies analog gain processes at RF and IF (at both pre- and post-log amplification) as well as processes in the digital domain. Intimately connected with thresholding strategies are signal processing, such as FTC, pulse integration, and correlation processes, both pulse-to-pulse and scan-to-scan. Modern digital technology, with its processing speed, available word length, and large memory capability, allows the radar designer to have great flexibility in the strategy applied, and what are entirely separate processes in the analog world increasingly become an integrated digital process. Simple thresholding ideas are being replaced by complex logical processes, making a detailed evaluation of whether a potential target is present or not, even potentially merging into plot extraction and tracking processes.

More complex processing can also yield additional information that is useful to marine operators. For instance, by applying spectrum analysis techniques, it is possible to extract accurate sea-state information, including significant wave height and period, direction, and speed.[16] The Miros A/S Wavex™ system determines directional wave spectra scaled in m^2/Hz and parameters such as significant wave height and average wave period. An FFT is performed on data collected on a scan-by-scan basis; 32 scans of data are typically used in the analysis. The resultant information can be highly useful for large high-speed craft, perhaps traveling at up to 60 knots or more. It is also potentially useful for vulnerable vessels such as chemical carriers to ensure that appropriate action in heavy weather can be taken, particularly at night or in poor visibility. Ocean oil rig operations can also benefit from such systems. Information can be displayed to the user in the form of digital readouts of the primary parameters, as well as graphically displayed data. Installed systems normally extract the raw data from the existing 9 GHz radar and perform wave processing and display functions on a separate processor/display system. Spectrum analysis of this type can also detect oil slicks[17] because they reduce the amplitude of sea surface capillary waves. Such systems can be valuable on vessels assisting with clean-up operations and also for early detection of spills from oil rig operations.

It has been proposed[18] that additional processing of the cross-polar content of the received radar signal may be of benefit to marine operations where sea ice is a hazard. Due to structural changes that occur in older ice that affect the reflection of radar energy, it is possible to differentiate between potentially dangerous old ice, including glacial ice (icebergs), and single-season sea ice, which is normally less dangerous to navigation. This is because the cross-polar component of the reflection is significantly higher for old ice compared to new ice. The problem is how to affordably determine the cross-polar component.

In practice, it has been found that excellent ice detection can be made by optimized processing of conventional marine radar signals.[7] Because of the slow-moving nature of the targets of interest, averaging the radar image over many antenna scans using an optimized infinite impulse response (IIR) filter can give a very detailed image that allows the user to differentiate between ice and water areas. In particular, small ice features such as bergy bits (icebergs with dimensions above the waterline of less than 5 meters in breadth and greater than 1 meter in height) and growlers (less than 2 meters in breadth and less than 1 meter showing above the water line) are readily detected by such systems. Integration over 128 seconds has been found to be suitable, with the antenna rotating at 120 rpm. The increased capability of (ice) target detection in sea clutter with higher antenna rotation rates has also been demonstrated. Dramatic images of ice hazards can be produced by such radars, such as that illustrated in Figure 22.4.

FIGURE 22.4 Ice feature detection using IIR filtering on Rutter Sigma 6 radar processor (*Courtesy of Transport Canada*)

Solid-state CMR. Several factors have come together to promote the introduction of civil marine radars with solid-state transmitters. The most important of these is that affordable magnetron-based radars do not meet user demands when operating in heavy sea and precipitation clutter. Small craft and buoys become invisible on the display, creating danger to life. IMO recognized this problem and, in order not to constrain opportunities for innovative radar design, removed the requirement that 3 GHz radars be compatible with existing racons.

Gallium Nitride and other microwave power semiconductors,[19] developed primarily for broadband communications links, have enabled CMR manufacturers to use these in the radar transmitter to replace magnetron-based designs. Pulse compression techniques are used to reduce the required peak power. Even single Gallium Nitride devices can generate hundreds of watts of peak power, at mean powers easily sufficient for CMR applications. Also, advances in digitally controlled waveform generators have given designers the ability to create pulse-compressed waveforms with high precision and at low cost. These waveforms enable coherent processing of the received signal, giving additional doppler information that can be used to help separate targets from clutter. The use of frequency diversity techniques to give added target detection possibilities becomes potentially affordable, because of the flexibility of the signal generation technology. Cost has precluded the use of dual-magnetron transmitters for this purpose.

Demands from other services for more bandwidth, particularly from mobile communications operators, continue to put pressure on the ITU-determined marine radar spectrum limits (see, for example, Williams[20]). Since the peak transmitted power from solid-state CMRs is very low compared to magnetron-based radars, for example 200 W compared to 30 kW, the spectrum interference levels are much reduced and, therefore, extended use of this technology could result in better use of the RF spectrum. Also, the highly controlled waveforms are expected to create less spectral noise than typical magnetron-based CMR transmitters.

An example of a solid-state coherent radar is the Kelvin Hughes SharpEye™—the first introduction of such an IMO-compliant system to the CMR market. It has a peak output power of 170 W and a duty cycle of 10%. Figure 22.5 shows a photograph of the transmitter electronics. In order to obtain the required short-range performance, it transmits a frame of pulses with differing lengths. Each pulse within the frame is optimized to cover a specified range bracket. Overall, the pulse sequence completely covers the instrumented range and ensures that the IMO specified minimum range requirement is met.

In the receiver, frames are grouped into blocks called bursts. The duration of a burst is approximately equal to the time taken for the 3 dB points of the antenna azimuth beam to sweep past a point target; consequently, the number of pulses in a burst is directly related to the instrumented range and the antenna rotation rate. The echoes received during a burst are processed by a filter bank to extract the radial velocities of targets and clutter. Within the digital signal processor, detection thresholds for each of the filters within the bank are calculated adaptively, aiming at providing optimum control of false alarms while maximizing clutter suppression and target detection. Manual control of the thresholds is also provided to be compliant with IMO requirements.

Modern, fully solid-state radars need little analog circuitry within their designs; they operate on low voltages and have no time-lifed components such as magnetrons. This potentially makes them extremely stable and reliable, with a resultant low cost of ownership, therefore, meeting the increasing demands of ship operators.

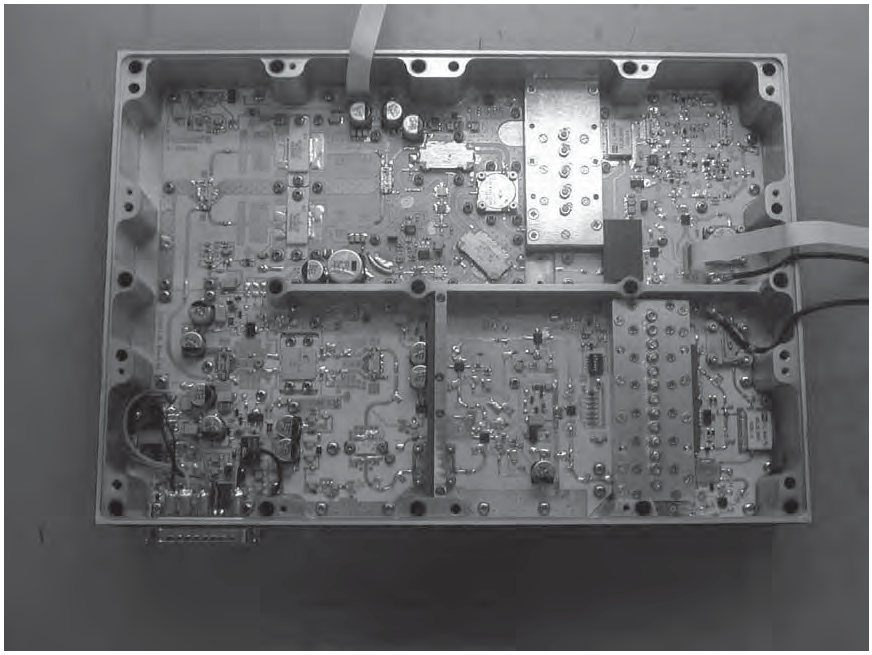


FIGURE 22.5 Kelvin Hughes SharpEye™ CMR S-band solid-state transmitter (*Courtesy of Kelvin Hughes Ltd.*)

In previous years, ships were required to have a radio officer, who could carry out radar repairs at sea. This is no longer the case. Reliability is a prime concern, as a non-operating radar can force the delay of the ship in port, at great cost to the operator.

22.5 TARGET TRACKING

The target tracking function of a shipborne navigation radar has historically been called an *Automatic Radar Plotting Aid (ARPA)*. This term is becoming obsolete. IMO now defines this process as *Target Tracking (TT)*, which includes target data obtained from AIS. The basic requirement calls for a minimum radar tracking capacity of 20 targets on ships less than 500 gt; 30 targets on ships between 500 and 10,000 gt; and 40 targets on ships over 10,000 gt. In addition, ships over 10,000 gt must have an automatic target acquisition capability. Actual systems commonly exceed these minimum requirements. Targets with a maximum relative speed of 100 kt must be trackable; this requirement is increased to 140 kt for radars on vessels capable of more than 30 kt. On the bridge, the navigator's requirements to aid collision avoidance include the need to know a target's *closest point of approach* (CPA) and *time to closest point of approach* (TCPA), both of which must be available for all tracked targets. The required tracked target accuracy at 95% levels is given in Table 22.4.

TABLE 22.4 Requirements for radar tracked target accuracy (95% levels)[11] (*Courtesy of IMO*)

Time of Steady State (minutes)	Relative Course (degrees)	Relative Speed (kn)	CPA (NM)	TCPA (minutes)	True Course (degrees)	True Speed (kn)
1 min: trend†	11	1.5 or 10% (whichever is greater)	1.0	–	–	–
3 min: motion‡	3	0.8 or 1% (whichever is greater)	0.3	0.5	5	0.5 or 1% (whichever is greater)

† Trend is an early indication (after 1 minute) of the target's speed and direction.
‡ Motion is the established assessment (after 3 minutes) of the target's speed and direction.

The tracking problem is complicated by the fact that basic radar measurements are made relative to the ship's motion, but the display may be set to relative or true motion. In addition, true motion can be ground or sea stabilized. Target vectors and associated data boxes can be shown in true or relative motion, whatever the frame reference of the radar display. A ship's orientation to north is given by a gyrocompass or "transmitting" magnetic compass—a compass with a digital interface. A log gives *speed through the water* (STW). This can be either a conventional rotating transducer driven by the movement through water or else an acoustic transducer measuring the doppler of the reflected signal. The latter can be set to assess speed either relative to the surrounding water (STW) or relative to the seabed, i.e., *speed over ground* (SOG). Mandatory carriage of a dual axis log (which measures speed in the forward and transverse directions) is required on ships above 50,000 gt. This is typically a doppler log. On smaller ships, GNSS is used to provide SOG and is often the user-preferred ground stabilization source for radar, even for a ship fitted with a doppler log. Doppler logs do not always give good speed readings on some types of seabed, for example, soft mud. Facilities must also be provided to allow the use of stationary tracked targets, such as radar conspicuous navigation marks, to provide the ground reference.

According to the design, the basic tracking function can be carried out in ship or ground/sea referenced frames, using conventional algorithms. The tracking process can be initiated manually or automatically. Automatic initiation is by a conventional plot extraction process confined within a user-defined area, which, at its simplest, could be a chosen range encircling the vessel. The defined area may also have user-defined exclusion zones. Algorithms for preventing plots from being formed on typically encountered wave features, perhaps lasting a few scans such as a traveling wave crest, need to be employed. Manual selection is effectively a plot extraction process operating over a small area surrounding the cursor. An alpha-beta tracker or other filtering technique is used to smooth measurement noise. The characteristics of this filter need to adapt to the quality of the received target signal. If tracking is carried out in ground referenced coordinates, the process automatically takes into account own-ship movements. In relative motion-based tracking systems, the filter needs to be aided with own-ship data.

Depending on the user-set mode, the data has to be converted to the correct reference frame and displayed appropriately. CPA and TCPA are continually calculated for all tracked targets, such that if limits preset by the user are breached, an alarm

can be initiated. All tracked targets are displayed on the screen with their associated velocity vectors. Tracked targets may be selected by the user such that all information concerning that target, including CPA and TCPA, is displayed on the data panel of the radar screen. Lost targets create a visual and an audible alarm. Normal termination of tracking occurs when a target leaves the acquisition zone or when manually cancelled. A guard zone may also be set up by the operator. This may be identical to the acquisition zone but is there to provide an alarm if any tracked target passes into the zone.

In common with other radar trackers, strategies have to be evolved to cope with the target being potentially invisible in some scans. IMO requires that the specified performance is maintained when the target is invisible in up to 50% of scans. Also, for an effective system, strategies have to be evolved to reduce the possibilities of target identities being swapped, which can happen when targets move close together and subsequently diverge. In particular, tracking algorithms have to attempt to cope with the potentially large and fast change in the radar centroid as a target vessel turns. In the worst case, this can amount to almost the length of the vessel, 300 meters or so, for a large ship. It is an art to get a good tracker optimized for all situations, over a variety of vessel speeds, and to maintain an appropriate indication of change in heading without excess latency. Over-damped systems may give an apparently stable indication of the track of a target but can be very inaccurate when a target changes heading. From the point of view of safety of navigation, the change in heading is often the more important parameter. Target trackers from individual manufacturers can have quite different design and optimization strategies and can, therefore, differ in performance. Within IEC 62388,[11] there are defined test scenarios that all SOLAS-approved tracking systems must meet. IMO requires that the trend in a target's change of direction is shown within one minute and the prediction of the target's motion should be available within three minutes, as given in Table 22.4.

In principle, target tracking could be aided by data from AIS (Section 22.7). However, AIS data is best left out of the radar tracking process in order to keep them entirely independent. Once radar tracks have been formed, they can then be automatically compared to AIS data and associated into a single track, if desired by the operator. This gives complete independence to radar- and AIS-derived data, therefore, enhancing integrity checking.

22.6 USER INTERFACE

From the user's point of view, the most visible and important change in marine radar from its early days has been the development of processor-based display technology. In particular, modern well-designed displays are viewable over a wide variation of ambient lighting; they make effective use of color and give easy and clear access to the radar image and associated data. The days when the radar screen was only viewable in daylight on dim long-persistence monochrome CRTs through the aperture of a hood are long gone. More recently, high brightness color CRTs are being replaced by Liquid Crystal Display flat panel technology, which is helping to make the display more user accessible—large stand-alone radar consoles are no longer necessary, allowing improvements to the ergonomic layout of a ship's bridge.

User input devices vary by manufacturer. Some solutions rely on little more than a tracker ball and three control buttons. Others have a number of dedicated switches and rotary controls, as well as a cursor control such as a tracker ball or joystick.

Touchscreen technology is sometimes employed. Increasingly, systems also include a full alphanumeric keyboard to allow easy input of user-supplied data, especially for *chart radars* (radars utilizing an electronic chart-based underlay) and radars integrated with AIS. Radars designed for smaller vessels tend to have a completely waterproofed user interface as they are often used in more exposed areas and by operators with wet, salty hands. In general, tracker balls, although they give more precise control and are common on shipborne radars, have been found to be unsatisfactory in the environmental conditions found on small craft. Instead, mini joysticks or simple four-way rocker switches are typically used.

The operational area of a radar display is normally circular, although this is no longer a mandatory requirement. It originates from the historical use of conical display tubes but is retained by most manufacturers as it gives additional space outside of the operational area for the display of data and menus (see Figure 22.6). The minimum operational display area is defined as a diameter: 180 mm for ships less than 500 gt; 250 mm for ships from 500 to 10,000 gt; and 320 mm for ships above 10,000 gt. The minimum recommended display areas for small craft radar are given in Table 22.2. The color of radar targets and background is not mandated. The target trails that used to be provided by the designed-in persistence of original radar monochrome CRTs have to be provided electronically. The trail length is required to be user selectable in units of time. When True Motion is selected, trails can be chosen by the operator to be shown in either true or ship-relative reference frames. The position of the display cursor is always available in a data box in terms of range and bearing from own ship and/or latitude and longitude coordinates. It is this cursor that is used to select and deselect targets within the operational display area and to draw user-defined maps. It is also commonly used to set range and bearing markers.

A *consistent common reference point* (CCRP) to which all radar and other navigational data need to be referenced is identified on the ship. This point clearly becomes of major importance when close-in navigational calculations are being made. Having a defined CCRP also allows a scaled own-ship's symbol to be shown on the radar display when appropriately short-range scales are selected. The symbology of this graphic, together with all other symbols and abbreviations on the display, should meet IMO requirements.[21] This ensures that operators are familiar with the radar presentation when working on different ships. The radar display should also comply with IMO's performance standards for navigational displays.[22]

Certain range scales (maximum displayed ranges) are mandatory, covering 0.25 to 24 nm. In practice, range scales above 24 nm are normally provided, typically up to 96 nm. Range rings may be optionally switched in by the operator to help estimate distances. Precise range measurements are made with the use of a *variable range marker* (VRM). At least two VRMs are needed, each with numerical readout in the data area of the display. An accuracy of 1% is required (but not better than 30 meters). A bearing scale around the periphery of the operational display must be visible. This scale can help users determine the ship's direction from viewing the *heading line* (HL), which has to be shown on the display; only temporary extinguishing of the HL is permitted. In addition, a ship's heading is normally available within a data box outside of the operational area. The radar origin can be offset from the center of the operational area by the user; the bearing scale adjusts accordingly.

Two or more *electronic bearing lines* (EBLs) have to be provided with continuous numerical readout. Although these are normally centered on the ship (at the CCRP), they can also be offset to any position. Readouts relative to own ship's heading or true north can be set. The EBL origin can also be set such that it follows own-ship movement or is

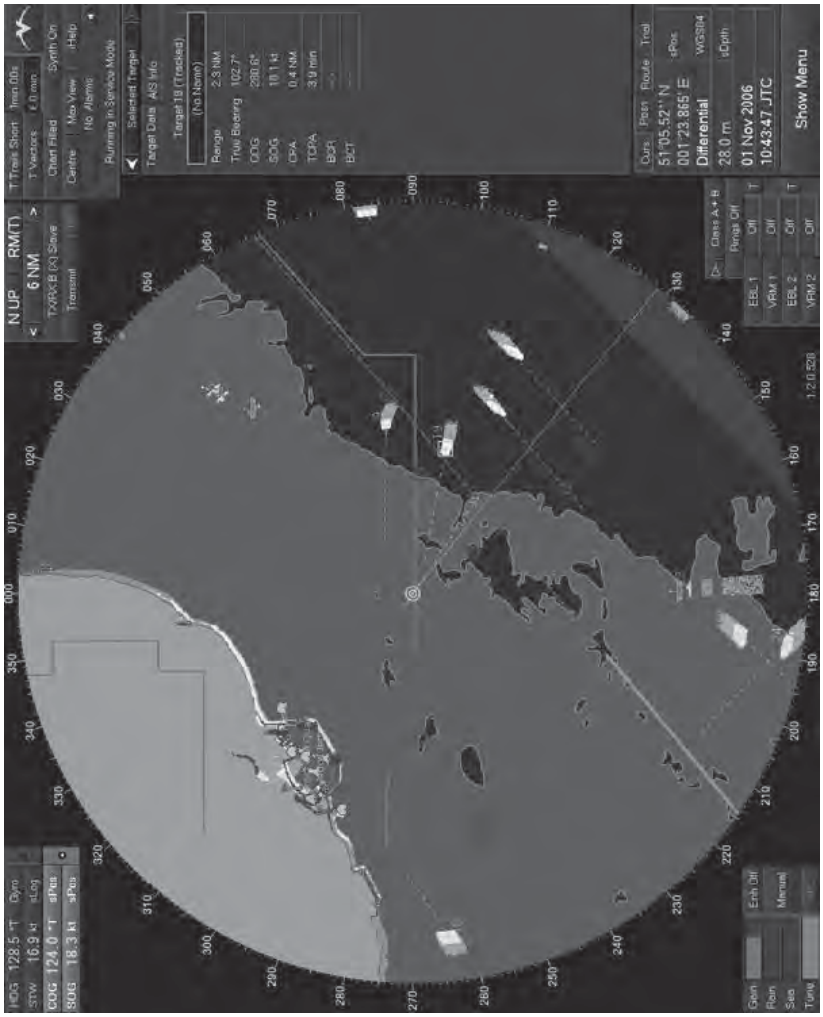


FIGURE 22.6 A ship's radar display (*Courtesy of Northrop Grumman Sperry Marine BV*)

geographically fixed. The distance and bearing of one point to another on the display can be determined, normally by use of a specific menu item and appropriate control of the display cursor. The actual implementation of EBLs, VRMs, and offset measurements is often effected by a common graphical tool, which is used to position and drag lines and circles across the display by means of the cursor. Many mariners find the use of ship-referenced parallel index (PI) lines to be highly useful. A PI is a straight line on the radar display that is user-set to a fixed "compass" bearing and a fixed perpendicular distance from the radar origin. At least four of these have to be provided. These can be individually switched into use and set by bearing, beam range, and length. PIs are typically used to ensure that a ship is maintaining a safe ground track, with reference to a ground-fixed conspicuous radar target.

Chart Radars. The capability and relatively low cost of modern processing and display systems allow great flexibility in the presentation of information to users. For many years, SOLAS-compliant radars have had the capability to employ user-defined maps as an underlay to the radar image. Many maps can be created and stored for future use. Although this facility is still widely used, the use of vectorized electronic chart data as a radar underlay is becoming more common. In IMO terms, they are known as *chart radars*. All approved chart radars have to be capable of displaying officially recognized vector data. This data is known as the *Electronic Navigational Chart* (ENC). It is issued on national authority and complies with an International Hydrographic Organization[23] (IHO) standard known as S57.[24] ENC data is normally displayed on approved electronic chart systems called *Electronic Chart Display and Information Systems* (ECDIS),[25] and may be used by ships in place of paper charts. ENC data is kept up-to-date by hydrographic offices that issue update files on a regular basis. An IMO-approved chart radar must also be able to accept these updates. Charts and their updates are loaded by CD-ROM or via a satellite communications link. On some systems, the chart radar may have access to a server on the ship that centralizes the distribution of such data to all equipment needing chart information.

The user has the ability to choose the ENC vector layers shown on the display of a chart radar. For instance, this may just include the coastline, navigation marks, and a single depth contour, considered safe for the draught of the ship. If the ship is navigating on ECDIS, rather than a paper chart, it is likely that both the radar and the ECDIS will be commonly set to Course-up or Head-up modes. North-up is no longer a particular advantage when the chart is not confined to such a presentation, such as paper charts. Most ECDIS equipment can optionally show radar-derived data, normally as tracked target vectors but sometimes as the radar image itself. This data is obtained from the radar processor via a digital interface, giving an apparent convergence of ECDIS and radar displays. Certainly, this is true at a basic design level, but IMO is keen to differentiate between the two. An ECDIS is used to plan and monitor passages; a radar is used primarily as a collision avoidance tool but also to aid position fixing, particularly by identifying ground-fixed radar conspicuous objects including coastlines. This results in many differences in the detailed requirements of radar and ECDIS displays. However, from a design point of view, the display processing requirements are very similar and can therefore use virtually identical hardware. As well as saving design costs, this enables an easy transition to *multifunction displays* (MFDs). An MFD can be instantly switched between radar and ECDIS, as well as other functions, enabling dynamic reconfiguration on a ship to optimize display use for particular circumstances. Clear indication of the selected mode becomes necessary for safety and statutory reasons.

Most small craft radars now being sold include the option of a chart underlay facility as a relatively low-cost option. In general, these use unofficial vector chart data, issued by specialist private companies. This data is more affordable than ENCs and is directed to this particular market. Because of cost and space constraints, a single display normally acts as both radar and electronic chart. Theses displays are all effectively MFDs and can, therefore, also be used as an electronic chart system without radar input.

22.7 INTEGRATION WITH AIS

The maritime Automatic Identification System[26] (AIS) is a target information system that performs similar functions to airborne Secondary Surveillance Radar (SSR), such as Air Traffic Control Beacon System (ATCRBS) and Identification Friend or Foe (IFF). However, the vast majority of transmissions are not the result of any interrogation, as it mainly operates as a broadcast system, based around a *Self Organizing Time Domain Multiple Access* (SOTDMA) communications protocol. The communications link, including the SOTDMA definition, is defined by the ITU.[27] Ships automatically transmit current navigational data and other information on VHF marine-band channels assigned for AIS use. The transmitted information is received by other ships and also by shore stations, such as coastal authorities and VTS facilities. Shore stations and ships also have the ability to specifically interrogate shipborne AIS transponders to initiate the sending of particular data. AIS has three major uses: to enhance the bridge team's situational awareness, to aid VTS activities, and to provide data to assist national security. The intention of IMO is that ships will normally display AIS data on the radar screen as it complements radar-derived data, adding to the integrity of the presence, position, and velocity of targets and also giving increased target information. In principle, a conventional secondary radar solution could have been adopted but international consensus favored the SOTDMA approach, as it was capable of providing higher levels of data exchange, particularly to aid VTS and security activities. An important advantage of the chosen AIS solution is its radio frequency. It is sufficiently low so that reasonable communications are maintained in situations where there is no visual or radar line-of-sight. This can be important in harbor, river, island, and estuary regions, where shielding by the terrain or buildings can affect radar range.

A shipborne AIS station broadcasts information divided into a number of sets. These comprise static data, such as the ship's name, type, length, and beam; dynamic data, including position, SOG, COG, and heading; and voyage-related data, such as destination port and ETA, depth under keel, and hazardous cargo type. The dynamic data is broadcast at a rate consistent with the vessel's velocity and whether it is changing course, as shown in Table 22.5. Static and voyage-related data is normally broadcast every 6 minutes. To provide sufficient bandwidth, two specific VHF 25 KHz channels are used, with stations alternating between channels at each message. There are 2,240 message slots per channel every minute. Minutes are aligned to Universal Time Coordinated (UTC), which is obtained from an integral GNSS receiver. The SOTDMA algorithm effectively reserves future slots for stations that are in reception range of each other, preventing mutual interference.

AIS for SOLAS use is known as *AIS Class A*. There is a Class B system that is designed for non-SOLAS use.[28] This system uses the same VHF channels as Class A, and the transmissions are necessarily compatible, but to avoid overloading the VHF data link (VDL), Class B uses Carrier Sensing TDMA. This is aimed at confining

TABLE 22.5 AIS Position Reporting Intervals (*Courtesy of IMO*)

Ship Dynamics	Reporting Interval (seconds)
At anchor or moored and not moving faster than 3 knots	180
With a speed of between 0–14 knots	10
With a speed of between 0–14 knots and changing course	3.33
With a speed of between 14–23 knots	6
With a speed of between 14–23 knots and changing course	2
With a speed of greater than 23 knots	2
With a speed of greater than 23 knots and changing course	2

Class B systems to use only slots unallocated to Class A users. Class B systems will delay their own transmissions if slots are not available. (There is an additional option of an SOTDMA-based Class B system.) Importantly, Class A and B systems receive each others' transmissions.

The combination of AIS and radar-derived data gives benefits to navigation because of the complementary nature of the two systems. The relative range and bearings of a target derived from AIS data are entirely independent of the radar measurements of these parameters. Clearly, any observed differences in radar and AIS positions will then indicate an error in some process, provided the differences are outside the expected noise in the measurements. This can be highlighted for the user. A high positional correlation increases the integrity of the observation, particularly as speed and course measurements can also be used in the comparison. Lack of any correlation can also give the user information that may be helpful. If only radar data is received, it may be that the target is not fitted with AIS, which means it could be a small craft, floating debris, or ice. It could also mean that a vessel's AIS is not operating or is transmitting erroneous positional information. If only AIS data is received, the radar image may be obscured by clutter, a headland, or even a poorly setup or faulty radar installation. Normally, just a few targets will be uncorrelated, highlighting that these few should be given additional caution if they are significant to own-ship navigation, at least until they can be positively identified, perhaps visually. If no targets are correlated, it suggests that own ship has a significant problem, perhaps with its radar, GNSS position, or more commonly, a gyrocompass offset.

If there is good positional tie-up or even an understanding of why radar information may be lacking, for example, due to heavy sea clutter, the additional information transmitted on AIS can be extremely useful. For instance, a target's heading is transmitted by AIS. This information is not available from radar (only the course can be determined), and yet heading is used as the basis for determining collision avoidance action. The AIS transmitted heading should align with the target's visual aspect and, therefore, the navigational lights on a vessel. Vessel names can be added automatically to target tracks on the radar/AIS display, and if there is a need to communicate on VHF with a particular target, the radio call sign is also available from the AIS data. The destination port and ETA can sometimes be useful in determining the likely intentions of targets, although such assumptions must be treated warily.

A significant advantage of radar is that it does not need cooperative targets. It attempts to detect all objects of potential interest. Its inherent Relative Motion mode of operation makes it ideally suited for collision avoidance use—particularly as in this mode it has no requirement to need own ship's geographical position. However, radar is basically confined to line-of-sight operation; its performance can be significantly

degraded by clutter, and its tracking capability is compromised when targets are changing course or passing close to other targets. AIS has reasonably good capability in non-line-of-sight situations because of its lower frequency. It quickly reports a target's change in heading or course, including rate of turn data, if it is available on the target vessel. AIS is not affected by sea clutter and can report absolute position accurately—to normally better than 10 meters or even a meter or two, if reporting differential GNSS-derived positional data. However, AIS relies on cooperative targets; is prone to gross errors in data accuracy, mainly caused by setup errors; and totally relies on reasonably accurate GNSS data being available. A GNSS blackout, perhaps caused by intentional or unintentional jamming, would prevent AIS from being an effective system, possibly over a wide area and for an appreciable time.

Future systems might increasingly exploit the complementary aspects of radar and AIS. This could improve the overall target-tracking capability available to ships and will give support for detecting cooperative targets in clutter. Possibly, the knowledge from AIS that a target is likely to be at a particular range and azimuth could direct concentrated processing techniques in that area, perhaps using pattern-matching algorithms as well as optimizing the false-alarm rate in the immediate area of the AIS-reported target. It is ironic that the very targets that may be hidden in clutter are smaller vessels that do not mandatorily carry AIS. Also, a Class B system only transmits once every 30 seconds at a maximum and so will be less useful for aiding radar even though it can usefully alert navigators that a small target is present.

The use of AIS as an *aid-to-navigation* (AtoN) has been put forward as a possible replacement for racons, which are described in the next section. In principle, AIS AtoNs could replace racons, but in practice, it would be a retrograde step as they cannot be used independently of a position fix system, such as GNSS. However, they can be usefully employed to indicate the integrity of the actual position of the mark, which may have dragged or become unattached, and other additional data, such as sea currents. The AIS transceiver does not have to be situated on the actual mark and could be shore-based to ease maintenance. Used in this way, they are known as Virtual AIS AtoNs.[29] Up-to-date information concerning the mark and its integrity can be automatically or manually fed-in by port authorities. Such systems can also be used for alerting mariners of the position of recent wrecks and other temporary and perhaps visually unmarked navigation hazards.

22.8 RADAR BEACONS

Radar beacons have played an important role in marine navigation ever since the early days of radar. They basically detect incident pulses from marine radars and instantaneously transmit a distinctive signal that identifies the beacon and its position on a radar display. There are three main uses of such beacons. The first of these is for enhancing visual aids to navigation, such as buoys and landmarks, to enable them to be prominently identified on a radar display. These are normally called *racons* (as a contraction of *radar beacons*). Such systems form an important navigational service that is well liked by mariners. The second use is for *Search and Rescue Transponders* (SARTs), which are mainly designed to be deployed from life rafts after a marine accident. The third category of use is for radar enhancement of small targets, such as pleasure craft. These are called *radar target enhancers* (RTE) or active radar reflectors.

Racons. The International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA) sets the performance standards for racons.[30] These incorporate the technical characteristics set out in a specific ITU-R Recommendation.[31] Because racons normally form only one subsystem of an AtoN. they therefore need to be small in size and power efficient since they are rarely connected to a main power supply. They often operate in an extreme environment, such as on a buoy buffeted by the sea. Racons are specified to meet an extended operational temperature range of –40° to +70°C. Modern racons operate by detecting an incident pulse and then measuring its frequency and responding at the same frequency, thereby reducing the interference potential with other in-band radars. They are often dual-band (3 and 9 GHz). The ITU recommends that for pulse lengths of 0.2 μs or longer, the frequency accuracy of the responding signal should be within ±1.5 MHz, and for pulses of less than 0.2 μs, the frequency should be within ±3.5 MHz. Swept frequency racons are effectively obsolescent but are still permitted. These work by having an internal RF source that is slewing in frequency across the entire radar band with a saw-tooth waveform at a rate between 60 and 120 seconds per 200 MHz. All received pulses are responded to but the interrogating radar will receive a racon burst only once every one to two minutes when its receiver is in band to the particular transmission.

IALA recommends that racons have suppression techniques to avoid responding to radar sidelobe transmissions. This is not an easy task to implement and probably impossible to make infallible. Basically, the racon needs to build a table of radar signatures that it is currently receiving, based on frequency and pulse length. It then identifies whether high level and lower level pulses of the same signature are being received and makes the assumption that these are from the same radar. It sets a threshold level for individual radars such that it only responds to high-power main beam interrogations. Typically, peak transmit powers are about 1–2 watts. Antennas are usually omnidirectional in azimuth but can have a restricted elevation beamwidth and typically have an overall gain of about 6 dB. Prime power consumption in average traffic can be less than 1 watt.

The modulation on the response signal of a racon paints a Morse code image on the radar display. The code identifies a particular AtoN and appears in the radial direction, conventionally commencing with a dash. This dash starts a short distance beyond the actual position of the AtoN because of inherent delays in the response time of the racon. However, delays giving an error of less than 100 meters are readily achievable. In good conditions, the AtoN primary radar image will be displayed on the radar screen, helped if a passive radar reflector is also a part of the AtoN. Racons have to include muting periods to allow ship radars to look for small targets in the vicinity of the racon identifier.

The long-term future of racons is unsure, although maritime authorities are assessing the situation.[32] Mariners like them as they are useful, familiar, and give ship-relative data. However, it is difficult to see how they will survive in their original form as marine radar moves away from utilizing magnetron-based systems. Also, compared with earlier years, when racons were essential, many more navigational aids are becoming available that assist positioning. These include multiple GNSS services, differential GNSS, AIS, and enhanced VTS facilities. There are also great improvements in onboard navigational aids such as electronic charts and integrated navigation systems.

The reliance of navigation on a single system, such as GPS, or even a single technology, such as GNSS, is not acceptable to the maritime community and neither to aviation. For instance, it is easy to jam all GNSS users over a wide area because of the small amplitude of the received signal. This means that radar and other positioning systems are likely to be always used as essential navigational tools. The overall

requirements for the electronic navigation of ships, including VTS reporting systems, are being examined by IMO and IALA, with the intention of determining a future e-Navigation concept (*e* for *electronic*/*enhanced*). The continued need for racons or a replacement technology will inevitably form part of this program. If the continuous availability of precise positional information cannot be totally relied upon, then it is probably essential that some form of ship-relative system to identify fixed navigational marks is available.

SARTs. Search and Rescue Transponders[33] (SART) form part of IMO's Global Maritime Distress and Safety System[34] (GMDSS). These are 9 GHz radar transponders that are mainly designed to be used on survival craft (such as life rafts) in emergency conditions. They are relatively small and affordable. On being triggered by a radar pulse, a SART emits a 12-cycle frequency-swept saw-tooth waveform covering 9.2 to 9.5 GHz. The extension down to 9.2 GHz covers the band used by search aircraft. The very fast upward frequency scan is accomplished in 0.4 μs; the downward scan takes 7.5 μs. This forms the possibility of a displayed trace on the radar screen, consisting of 12 radial dots and dashes as the upward and downward scans cross the passband of the radar receiver, with the first dot at a slightly longer range than the SART position. In practice, the upward sweep is so fast that the dots are normally not visible on the display and only the dashes can be seen. Even these can be quite difficult to locate in adverse sea clutter conditions.

The first dash displayed on the radar screen could be up to 0.8 nm away from the actual SART position and so search craft have to take precautions not to run down the survival craft when bearing down on the signal. At short ranges, the swept gain of the radar may truncate the nearer dashes. Also at short ranges, because there is no sidelobe suppression circuitry, SARTs can be triggered by radar sidelobes. To prevent adjacent SARTs from continuously triggering each other, there is a short delay after a SART transmission before it may be triggered again. To detect SARTs in heavy sea clutter, it is often best to detune the radar receiver, eliminating all other returns. Some radar manufacturers provide a SART search mode that sets the radar optimally for their detection, including inhibiting pulse-to-pulse correlation and optimizing filter bandwidths. AIS-based SARTs have now been proposed. These may eventually replace radar-based SARTs because the latter are difficult to detect in adverse conditions.

Radar Target Enhancers. Radar target enhancers[35] (RTEs) are used increasingly by small craft because, for their size, they offer a better enhancement in radar cross section than can be given by a passive reflector. In principle, they are simple devices. In-band received signals are amplified and retransmitted with minimum delay. Delays can be kept to a few nanoseconds, less than the equivalent dimensions of the craft, ensuring co-located returns of the enhanced signal and natural radar reflection. To prevent positive feedback between receiver and transmitter, the transmit and receive antennas are normally physically separate, one above the other, providing isolation. Isolation can also be increased by transmitting on an orthogonal polarization to that received. Providing they are operating with linear gain, there are no adverse effects from radar sidelobe interrogation. However, at close ranges, the signal from the main beam of the radar may saturate within the RTE, effectively enhancing the levels of the RTE received signal through the radar sidelobes. The ITU regulations limit RTEs to an EIRP of 10 watts, with a minimum gain of 50 dB.

22.9 VALIDATION TESTING

The factors affecting the range performance of a radar system are well known, and increasingly sophisticated design methodologies have greatly improved the detection of all forms of radar. The final proof, however, is how the radar actually performs at sea. As stated previously, shipborne radars are validated as meeting IMO performance standards by being independently type approved to technical standards issued by the IEC. The IEC standards include defined methods of testing. For a given target and radar antenna height, it is relatively easy to define and execute a test to determine that a point source target with a specific echoing area is detected at a given range in a minimal clutter field. It is very difficult to extend this to determine, in a repeatable and quantitative manner, the performance of the radar against point targets in predefined clutter conditions. For this reason, some basic performance tests have necessarily been loosely defined to allow scope for approved test laboratories to make their own qualitative judgments on basic radar performance, normally based on opportunistically testing the radar over the sea and in precipitation in a variety of situations. Judgments on performance can, therefore, be quite subjective and are naturally affected by the conditions actually encountered during the tests. Cost considerations can severely limit the length of test programs and thereby the range of scenarios used. Radars under type approval are typically installed on a trial vessel for such tests or use a land site overlooking the sea.

This scenario is becoming increasingly unsatisfactory as advances in requirements for safety and the protection of the environment mean that it is necessary to ensure that type approval is consistently applied and is, therefore, measured in a quantitative manner. In an attempt to resolve this, some work has been performed to try to better formalize marine radar clutter performance tests, including some research performed on behalf of the UK Maritime and Coastguard Agency.[36] This approach was aimed to minimize any special configuration of the radar under test. It is based on a system that generates simulated target and clutter waveforms. These are picked up by the antenna of the radar under test from a nearby transmitting source, typically situated about 100 meters from the radar antenna. Co-located with the transmitting source is a receiver, which detects the transmitted radar signal and continually analyses its frequency, pulse length, and amplitude as the radar antenna rotates. From this information, a signal waveform is synthesized on a pulse-to-pulse basis, replicating reflected signals from targets and clutter. The synthesis procedure calculates appropriate fluctuating target and clutter returns from any desired theoretical model—which could also, in principle, include models derived from recorded data of real target and clutter reflections. Because the simulated signal is predominately entering the radar through the sidelobes of the radar antenna—except when the radar main beam aligns with the simulator antenna—the synthesized signal needs to be automatically adjusted in amplitude to compensate for the actual sidelobe sensitivity in the direction of the simulator. Effectively, the synthesizer has to amplify the transmitted signal according to the inverse of the amplitude of each pulse received from the radar. The challenges in designing an affordable system include the large dynamic ranges that have to be encompassed and the processing speed needed to determine the characteristics of the transmitted signal.

In principle, a number of clutter and target models could be established by international agreement, such that they were considered to be representative of conditions around the world; agreed-upon test criteria could then be determined, and simulator systems could be based at marine radar–type approval laboratories. It has been found

that the system is adversely affected if set up close to large radar-reflecting objects, such as buildings, and so needs to be located ideally in relatively open spaces, similar to that found on an antenna far-field test range. To enable tests to be more easily carried out, direct injection of test targets and clutter into the RF path is feasible. However, this involves a certain amount of adaptation of the radar under test, which may be considered inappropriate.

The potential move at 3 GHz to pulse compressed radar offers additional challenges in the design of a universal simulator, as the system described is based on testing noncoherent pulsed radars. Systems based on digital RF memory may have to be devised, storing waveforms that can be subsequently processed. Target and clutter models would obviously need to appropriately take into account doppler effects introduced by the movement of their equivalent scatterers.

22.10 VESSEL TRACKING SERVICES

Radar heads for port control and coastal surveillance systems have some requirements in common with those for shipborne radars. This originally resulted in many of the well-known suppliers of shipborne radar getting involved in this area. They could offer attractive prices as the subsystems were derivatives of the relatively high volume shipborne market. Over time, much of the market has become more sophisticated, and because of this, specialist organizations now dominate the supply of systems for this application. The large costs associated with a major vessel tracking services (VTS) operation, including massive antenna support towers, operations buildings, specialized software, and disaster-proofed broadband communication systems, mean that the costs of a more optimized radar head often become a relatively insignificant addition. This also means that switchable linear and circular polarization modes are more common on VTS systems. However, basic low-cost VTS systems still commonly use subsystems intended for shipborne use, giving good cost savings compared to custom-made systems.

There are significant differences, however, in the requirements for a VTS radar compared to a shipborne system. The VTS antenna is mounted on a static platform. This means that the vertical pattern can be more optimally shaped. Also, since the design does not have to cope with the shock, vibration and instability experienced on ships' radar masts, larger antennas become feasible. This allows azimuth beamwidths to be narrower, therefore, reducing the size of clutter cells. The required coastal area to be covered can be large, and getting the best range out of a few radar heads situated on tall towers is often more cost effective than utilizing many smaller installations. Because VTS often forms part of a nation's security network, then a longer range capability than that just required for port operations may be necessary. This implies that very high antenna towers are often needed, in some cases up to 100 meters. This exacerbates vertical lobing effects, which may need to be reduced by the use of vertical pattern shaping. The long-range requirement often means that greater transmitter power than that used on shipborne radars is needed, even though VTS antenna gains can be higher. Pulse lengths must be kept short to get good clutter immunity, but simultaneously, long-range performance is required, again increasing the required transmitted power. VTS radar heads are usually not operator configurable because a number of operators can typically be using data from one head.

There are more opportunities to enhance performance because of the fixed antenna position; for instance, sea clutter mapping becomes easier because the antenna is not on a moving platform. Also, the clutter conditions can be less variable because of the restricted geographical area of operation, and there are no degradations in the accuracy of the displayed radar image in having to compensate for a ship's heading with compass input. In particular, target tracking is performed from a stable and static platform. However, it is generally necessary to track many more targets than is required on a shipborne radar, and normally VTS has fully automatic plot extraction and track initiation. Also, more information on tracked targets may need to be easily available. Much of this additional data can be automatically supplied by AIS. The radar data often has to be relayed many miles, to perhaps a number of operational centers. It may need to be combined with data from a number of radar heads and, therefore, will be quite synthetic when displayed on operators' screens, reducing the possibilities of individual operator adjustment. Extensive data communications networks become a critical aspect in the performance of the VTS. High reliability of the system is required because of safety, environmental protection, and security aspects. A total system availability of 99.9% is not uncommonly specified, implying an average downtime of less than 2 minutes per day.

Another major difference compared to shipborne radar is the custom nature of the installation. Radar heads are fixed, and there is a specific requirement for certain performance parameters to be met in the particular localized environment. Sea clutter, although very variable, will have certain local characteristics, enabling more effective optimization of the processing. In particular, the actual performance can be more easily measured against design requirements.

The design of high-performance antennas for VTS applications has a similarity to air traffic control antennas, in that they both ideally require a tailored elevation pattern. The ideal pattern shaping for a high-mounted VTS antenna requires a sharp cut-off above the horizon and a tapered pattern below. Energy directed above the horizon increases precipitation clutter and also reduces the antenna gain. At angles below the horizon, the gain should nominally follow a cosecant squared power law. This is aimed at giving a constant signal strength from a target of fixed RCS, independent of range. These are often known as *inverted* or *inverse* cosec squared antennas to differentiate them from air traffic control radar antennas that have their shaping at angles above the horizon. Such shaping optimizes the pattern to the application, greatly enhancing overall performance. Typically, the pattern shaping is enabled by a doubly curved reflector fed from a point-source primary feed. An example from Easat Antennas is illustrated in Figure 22.7. This is a 7.5 meter reflector antenna with a 35 db gain at 9.3 GHz. It has an inverse cosecant-squared elevation pattern and a 0.3° azimuth beamwidth. It is remotely controlled to give horizontal or circular polarization. The vertical pattern shaping interacts with the STC of the radar receiver, and it is, therefore, necessary to take this into account in the system design. To improve detection, frequency diversity is often used on prime systems.

IALA has issued detailed recommendations[37] on the operational and technical performance requirements for VTS equipment. There is much useful information in the recommendations, and they are essential for procurers and designers of VTS radar equipment. They cover both coastal and waterway installations. Many major rivers of the world carry vast amounts of cargo on ships that can be surprisingly large. The meandering nature of rivers and relatively abrupt turns in canal systems, together with the natural and manmade obstructions to radar, mean that waterway vessel traffic systems are generally covered by many low-power radar heads on relatively low towers.

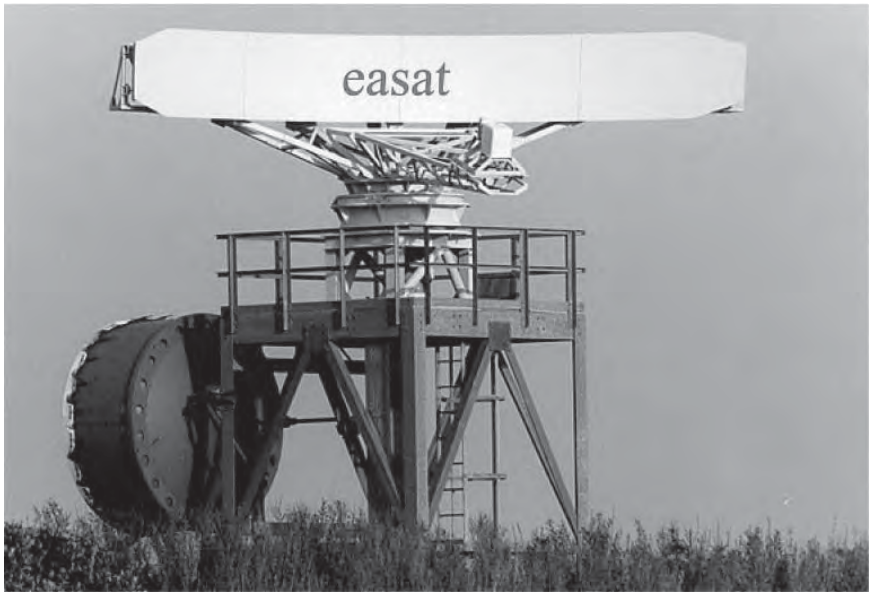


FIGURE 22.7 Dual-polarized doubly curved VTS radar antenna (*Courtesy of Easat Antennas Ltd.*)

Because of their number, such radars tend to be minimally adapted shipborne radars as they offer adequate performance at relatively low cost.

Interestingly, the IALA recommendations allow CCTV solutions to compete with radar when there is very low traffic density. Automatic tracking of a single target is then required. However, on basic radar-based systems, IALA expects a 100 target track capability and a plot extractor that can deal with more than 1,000 plots per rotation. On an advanced system, more that 300 targets may have to be tracked with a possibility of more than 5,000 plots per antenna revolution.

APPENDIX THE EARLY DAYS OF CMR

The use of commercial marine radar arose directly from the rapid development of radar technology for military applications during World War II. Even as early as 1944, some attention was being given to the peacetime role of radar as a navigational aid for commercial shipping. In 1946, an "International meeting on radio aids to marine navigation" was held in London and attended by representatives from 23 countries.[38] The meeting was chaired by Sir Robert Watson-Watt. It was seen that radar on commercial vessels had an important part to play in anti-collision, coastal navigation, and pilotage decisions. (*Pilotage* is navigating in waters where a qualified pilot is required to be onboard.) The future compulsory fitting of radar to ships was contemplated, as was the desirability of an internationally agreed upon minimum performance standard, with requirements for nationally issued certificates of type approval. The need to include the use of radar within the International Collision Regulations was clearly seen, together with the need for certification of users.

In 1946, the UK favored operation at 9 GHz, presumably as it more affordably met the UK perceived azimuth requirements of 3° resolution and 1° accuracy. The United States identified operational problems at 9 GHz that could be experienced in the extreme rainfall conditions found on the U.S. eastern seaboard. These caused "blackouts" on early 9 GHz systems—defined as an effective range of less than 1 mile. As a consequence, the U.S. favored operation at 3 GHz. The shortest pulse lengths then commonly available (around 250 ns) made the clutter cells large, resulting in 9 GHz radars being very susceptible to rain clutter, particularly as clutter processing techniques were in their infancy. In 1946, there was no question that a commercial ship could afford both a 3 and a 9 GHz radar, as the expense of even a single radar system was seen to be a limiting issue. Because of cost, it was already envisaged that fitment would be confined mainly to certain classes of passenger ships that had a definite need to carry radar, particularly those working in the north Atlantic, in congested areas or areas subject to fog or ice.

The early trials in the UK concentrated on a single 9 GHz demonstration system fitted to a naval vessel. It was based around a 40 kW magnetron capable of 250 ns pulses at a PRF of 1,000 Hz. It was interesting that the speed of rotation could be varied between 20 and 100 rpm. Despite the perceived modernity of today's chart radars, it was connected to an optional *Chart Comparison Unit*, which was an optical system allowing the radar image to be displayed in coincidence with a paper chart. The facility to allow "North-up" operation was always seen to be a vital requirement for marine navigation radar. Parallel trials in the United States were conducted on a number of candidate systems, using a broad range of frequency bands. The initial trials were conducted in the Great Lakes and were overseen by the Coast Guard.

The radar standards proposed in 1946 were not adopted internationally, although the UK issued national performance standards based on them in 1948. The UK standard was also adopted by a number of other countries. It was not until 1971 that international marine radar standards were agreed by the Intergovernmental Maritime Consultative Organization (IMCO, the original name of IMO). However, the use of radar on ships was first formally recognized by IMO in 1960 in an Annex to the International Regulations for Preventing Collisions at Sea. The influence of the 1946 proposed international standard was evident in the 1971 performance standards, even to the extent of using identical wording in a number of places.

The similarity in performance requirements is still evident in the latest revisions of the IMO performance standard. For instance, the 1946 proposed performance specification included the need to give a clear indication of coastlines rising to 200 ft at 20 miles, of a 500 gross registered ton ("Imperial" units) vessel at 7 miles and of a 30 ft fishing vessel at 3 miles. The modern performance requirements, summarized in Table 22.1, still use these figures but with parameters, except ranges, given in equivalent metric units.

The technical vision of the 1946 meeting was remarkable. For instance, it was seen that in the future it would be possible to overlay radar data automatically onto a chart image displayed on a "television" type screen. This was not to be realized on commercial systems for 50 years. Also, it was observed that such display systems could accomplish more than one function and not just be used for showing radar on a chart. This anticipated the concept of multifunction displays, now in use on some integrated bridge systems.

It is interesting to note that Kelvin Hughes and Decca obtained the first type approval for commercial marine radar in 1948; effectively, both are still supplying

marine radar today. Kelvin Hughes has retained its name and Decca is incorporated into the Sperry Marine organization of Northrop Grumman Corporation. The 1948 Kelvin Hughes Type 1 radar had a peak power of 30 kW, 0.2 μs pulse width, and a PRF of 1,000 Hz. The 5 ft (1.5 meter) cheese antenna had horizontal and vertical beamwidths of 1.6° and 11°, respectively, rotating at 30 rpm. There was an antenna heater to prevent icing, and the transmitter and the receiver (to IF) were "upmast" (integrated within the antenna turning unit.) The display was a 9 in (23 cm) cathode ray tube plan position indicator. The similarities with systems being sold in the 21st century are perhaps more surprising than the obvious differences.

LIST OF MARITIME RADAR-RELATED ABBREVIATIONS

AIS	Automatic Identification System
AtoN	Aid to Navigation
CCRP	Consistent Common Reference Point
CMR	Civil Marine Radar
COG	Course Over Ground
CPA	Closest Point of Approach
EBL	Electronic Bearing Line
ECDIS	Electronic Chart Display and Information System
ENC	Electronic Navigational Chart (The data for ECDIS)
FTC	Fast Time Constant (differentiator)
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
gt	Gross tonnage (metric tonnes)
HL	Heading Line
IALA	International Association of Lighthouse Authorities
IBS	Integrated Bridge System
IEC	International Electrotechnical Commission
IMO	International Maritime Organization
ITU	International Telecommunications Union
MFD	Multi-function Display
PI	Parallel Index line
NT Radar	New Technology Radar (Marine term for coherent solid-state radars)
nm	Nautical mile (= 1842 meters)
SART	Search and Rescue Transponder
SOG	Speed Over Ground
SOTDMA	Self Organizing Time Division Multiple Access
STW	Speed Through the Water
TCPA	Time to Closest Point of Approach
VTS	Vessel Traffic Services
UTC	Universal Time Coordinated
VRM	Variable Range Marker

ACKNOWLEDGMENTS

1. Material from IMO publications is reproduced with the kind permission of the International Maritime Organization (IMO), which does not accept responsibility for the correctness of the material as reproduced: in case of doubt, IMO's authentic text shall prevail.
2. The author thanks the International Electrotechnical Commission (IEC) for permission to reproduce information from its International Standard IEC 62252 ed.1.0 (2004). All such extracts are copyright of IEC, Geneva, Switzerland. All rights reserved. Further information on the IEC is available from www.iec.ch. IEC has no responsibility for the placement and context in which the extracts and contents are reproduced by the author, nor is IEC in any way responsible for the other content or accuracy therein.

REFERENCES

1. International Maritime Organization, www.imo.org.
2. "Revised recommendations on performance standards for radar equipment," Resolution MSC.192(79), International Maritime Organization, London, 2004.
3. International Association of Marine Aids to Navigation and Lighthouse Authorities, www.iala-aism.org.
4. "Technical characteristics of maritime radio-navigation radars," ITU-R Recommendation M.1313, International Telecommunication Union, Geneva.
5. International Telecommunication Union, www.itu.int.
6. J. Crony, "Civil marine radar," in *The Radar Handbook*, 1st Ed., M. I. Skolnik. (ed.), New York: McGraw-Hill, 1970, Chapter 31.
7. J. Ryan and C. Kirby, "Iceberg detection performance analysis," Report TP 14391E, Transportation Development Centre, Transport Canada, 2005.
8. J. N. Briggs, "Target detection by marine radar," Institution of Electrical Engineers (now the Institution of Engineering and Technology), London, 2004.
9. "The international convention for the safety of life at sea (SOLAS), 1974," International Maritime Organization, London, as amended.
10. International Electrotechnical Commission, www.iec.ch.
11. "Maritime navigation and radiocommunication equipment and systems—Shipborne radar," IEC 62388, International Electrotechnical Commission, Geneva, 2007.
12. "Maritime navigation and radiocommunication equipment and systems—General requirements," IEC 60945, International Electrotechnical Commission, Geneva, 2002.
13. "Maritime navigation and radiocommunication equipment and systems—Digital interfaces," IEC 61162 (series), International Electrotechnical Commission, Geneva.
14. "Regulations regarding the minimum requirements and test conditions for radar equipment used for River Rhine and inland waterways," Central Commission for the Navigation on the River Rhine, Strasbourg, 1989.
15. "Maritime navigation and radiocommunication equipment and systems—Radar for craft not in compliance with IMO SOLAS Chapter V," IEC 62252, International Electrotechnical Commission, Geneva, 2004.
16. R. Gangeskar and Ø. Grønlie, "Wave height measurements with a standard navigation ship radar, results from field trials," presented at Sixth International Conference on Remote Sensing for Marine and Coastal Environments, Charleston, South Carolina, 2000.

17. R. Gangeskar, "Automatic oil-spill detection by marine X-band radars," *Sea Technology*, August 2004.
18. T. K. Bhattacharya et al., "Cross-polarized radar processing," Report TP 13263E, Transportation Development Centre, Transport Canada, 1998.
19. R. Pengelly, "Improving the linearity and efficiency of RF power amplifiers," *High Frequency Electronics*, September 2002.
20. P. D. L. Williams, "Civil marine radar—a fresh look at transmitter spectral control and diversity operation," *The Journal of Navigation*, vol. 55, pp 405–418, 2002.
21. "Guidelines for the presentation of navigation-related symbols, terms and abbreviations," Safety of Navigation Circular 242, International Maritime Organization, London, 2004.
22. "Performance standards for the presentation of navigation-related information on shipborne navigational displays," Resolution MSC.191(79), International Maritime Organization, London, 2004.
23. International Hydographic Organization, www.iho.shom.fr.
24. "Transfer standards for digital hydrographic data," Publication S-57, International Hydrographic Organization, Monaco.
25. H. Hecht, B. Berking, G. Büttgenbach, M. Jonas, and L Alexander, *The Electronic Chart*, 2nd Ed, Lemmer, Netherlands: GITC, 2006.
26. "Operational use of AIS," Model Course 1.34, International Maritime Organization, London, 2006.
27. "The technical characteristics for a universal shipborne automatic identification system (AIS) using time division multiple access in the maritime mobile band," ITU Recommendation M.1371-1, International Telecommunication Union, Geneva.
28. "Maritime navigation and radiocommunication equipment and systems—Class B shipborne equipment of the automatic identification system (AIS)," IEC 62287-1, International Electrotechnical Commission, Geneva, 2006.
29. "Recommendation A-126 on the use of the automatic identification System (AIS) in marine aids to navigation," Edition 1, International Association of Lighthouse Authorities (IALA), Paris, 2003.
30. "Recommendation R-101 on marine radar beacons (racons)," Edition 2, International Association of Lighthouse Authorities (IALA), Paris, 2004.
31. "Technical parameters for radar beacons (racons)," ITU Recommendation M.824-2, International Telecommunication Union, Geneva.
32. A. P. Norris, "The future of racons," Final Report, Contract No 237293, General Lighthouse Authorities, London, 2006.
33. "Global maritime distress and safety system (GMDSS)—Part 1: Radar transponder—Marine search and rescue (SART)," IEC 61097-1, International Electrotechnical Commission, Geneva, 1992.
34. *GMDSS Handbook*, 2nd Ed., London: International Maritime Organization, 2000.
35. "Technical parameters for radar target enhancers" ITU Recommendation M.1176, International Telecommunication Union, Geneva.
36. T. P. Leonard and S. J. Brain, "Radar performance test Methods—final report," Research Project RP544, UK Maritime and Coastguard Agency, Southampton, 2005.
37. "Recommendation V-128 on operational and technical performance requirements for VTS Equipment," Edition 2.0, International Association of Lighthouse Authorities (IALA), Paris, 2005.
38. "International meeting on radio navigation aids to marine navigation, May 1946," vol. 1 *Record of the meeting and demonstrations*, His Majesty's Stationery Office, London, 1946.

Chapter 23
Bistatic Radar

Nicholas J. Willis

Technology Service Corporation (retired)

23.1 CONCEPT AND DEFINITIONS

A bistatic radar uses antennas at separate sites for transmission and reception. The transmitter and receiver can be, and usually are, located at those sites to minimize transmission line losses. In nearly all cases of bistatic operation, antenna separation is selected to achieve some operational, technical, or cost benefit, and is usually a significant fraction of the target range.[1] Bistatic radars have been designed, developed, tested, and in some cases, deployed for military, commercial, and scientific applications. Typical military applications include air and space surveillance and range instrumentation. Commercial applications include wind field measurements and traffic surveillance. Scientific applications include measurement of planetary surfaces and atmospheres and study of ionospheric turbulence. Examples are given in Section 23.4. While these examples are both credible and useful, they are niche applications when compared to the ubiquitous capabilities of monostatic radars, which remain the principal method for *radio detection and ranging*.

Bistatic radars can operate with *dedicated transmitters*, which are designed for bistatic operation and controlled by the bistatic radar, or with *transmitters-of-opportunity*, which are designed for other purposes but found suitable for bistatic operation even when not controlled by the bistatic radar. When the transmitter of opportunity is from a monostatic radar, the bistatic radar is often called a *hitchhiker*. When the transmitter of opportunity is from a broadcast station or communications link, sources other than a radar, the bistatic radar has been called many things including *passive radar, passive bistatic radar, passive coherent location, parasitic radar,* and *piggy-back radar.*[2] Transmitters-of-opportunity in military scenarios can be designated either *cooperative* or *noncooperative*, where cooperative denotes an allied or friendly transmitter and noncooperative denotes a hostile or neutral transmitter.

Bistatic target detection uses a process similar to that of a monostatic radar, where the target is illuminated by a transmitter and target echoes are received, detected, and processed by a receiver. When operating with transmitters using CW or high-duty cycle waveforms, a bistatic receiver may need to augment its spatial isolation with spatial and/or spectral cancellation to reduce the transmitter's *direct-path* feed-through to acceptable levels. The bistatic radar can also use a portion of the residual or uncancelled direct-path transmit signal as a reference in a *correlation receiver*, which cross-correlates the received and transmitted signals, emulating matched filter operation.

Bistatic target location uses a process different from that of a monostatic radar. In a typical implementation, the bistatic radar measures (a) the transmitter-to-target-to-receiver propagation time, converted to a transmitter-to-target plus target-to-receiver *range-sum*; (b) the target direction-of-arrival (DOA) from the receiver; and (c) the transmitter-to-receiver distance, or *baseline*, to solve the transmitter-target-receiver triangle, called the *bistatic triangle*. This triangle locates the target, usually in terms of a range and angle referenced to the receive site. Other location schemes are given in Section 23.6.

When separate transmit and receive antennas are at a single site, as is common in CW radars, the radar has characteristics of a monostatic radar, and the term *bistatic* is not used to describe such a system. In special cases, the antennas can be at separate sites, and the radar is still considered to operate monostatically. For example, an over-the-horizon (OTH) radar can have a site separation of 100 km or more to achieve adequate transmit signal isolation. But that separation is small compared to the target range of thousands of kilometers, and the radar again operates with monostatic characteristics.

A variation of the bistatic radar is the *multistatic radar*, which uses multiple antennas at separate locations, one antenna for transmission and multiple antennas—each at a different location—for reception, or vice versa. Again, transmitters or receivers are usually sited with the antennas. Target detection is done bistatically, with each transmit-receive pair performing independent detections within a surveillance region common to all such pairs. Target location typically measures the baseline and takes simultaneous range-sum measurements from multiple transmitter-receiver pairs, which are plotted as ellipses with a transmitter-receiver pair at each ellipse foci. The intersection of these ellipses, or *constant range-sum contours*, locate the target. It is similar to *multilateration* because only range measurements are used to locate the target.*

A multistatic radar can also use *triangulation* for target location by taking simultaneous target DOA measurements from multiple receive sites at known locations. It is used by *SPASUR*[12] as a brute-force, satellite location technique. However, because of the large aperture sizes (or array lengths) required for sufficiently accurate DOA measurements at useful ranges, triangulation is seldom considered for other applications.

Concepts, data, and expressions developed for bistatic radars often apply to multistatic radars, for example, the range equation, target doppler, target radar cross section, and surface clutter. Thus, the remainder of this chapter will concentrate on the bistatic radar topic, developing multistatic excursions and deviations only when necessary.

Passive receiving systems, or *electronic support measure* (*ESM*) systems, often use two or more receiving sites. Their purpose is typically to detect, identify, and locate transmitters such as those from monostatic radars. They are also called *emitter locators*. Target location is by means of combined angle measurements from each site (e.g., triangulation) or time-difference-of-arrival and/or differential doppler measurements between sites (e.g., multilateration). These systems usually are not designed to detect and process the echoes from targets illuminated by the transmitter. They can, however, be used by a bistatic hitchhiker to identify and locate a suitable transmitter. Thus, although they have many requirements and characteristics common to multistatic radars, they are not radars and will not be considered here.

---

* The foregoing describes multistatic operation that combines data noncoherently. Coherent data combining is also possible where for example in-phase and quadrature data from each receive site is combined to form a large receive aperture. Examples include thinned, random, distorted, and distributed arrays,[3–7] interferometric radars,[8,9] and the radio camera.[10,11] This subject is treated further in Willis.[1]

The foregoing definitions are broad and traditional[13–16] but are by no means uniformly established in the literature. Terms such as *quasi-bistatic, quasi-monostatic, pseudo-monostatic, tristatic, polystatic, real multistatic, multi-bistatic*, and *netted bistatic* have also been used.[17–20] They are usually special cases of the broad definitions given above. The term *pseudo-monostatic* will be used to characterize bistatic geometries that approximate monostatic operation.

23.2 COORDINATE SYSTEMS

A two-dimensional north-referenced coordinate system[21] is used throughout this chapter. Figure 23.1 shows the coordinate system and parameters defining bistatic radar operation in the (x, y) plane, also called the *bistatic plane*.[22] The bistatic triangle lies in the bistatic plane. The distance L between the transmitter and the receiver is called the *baseline range* or simply the *baseline*; R_T is the range between transmitter and target; and R_R is the range between receiver and target. The angles θ_T and θ_R are, respectively, the transmitter and receiver *look angles*, which are taken as positive when measured clockwise from north. They are also called direction-of-arrival (DOA), angle-of-arrival (AOA), or line-of-sight (LOS). The *bistatic angle*, $\beta = \theta_T - \theta_R$, is the angle between the transmitter and receiver with the vertex at the target. It is also called the *cut angle* or the *scattering angle*. It is convenient to use β in calculations of target-related parameters and θ_T and θ_R in calculations of transmitter- and receiver-related parameters.

The transmitter-to-target-to-receiver range measured by a bistatic radar is the range sum ($R_T + R_R$). Methods for measuring this sum are given in Section 23.6. The range sum locates the target somewhere on the surface of an ellipsoid whose foci are the transmitter and receiver sites. The intersection of the bistatic plane and this ellipsoid

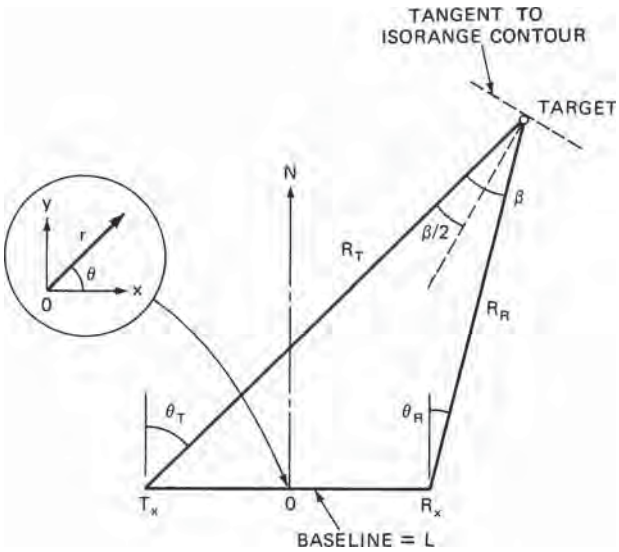


FIGURE 23.1 Bistatic radar north coordinate system for two dimensions, establishing the bistatic plane. The bistatic triangle lies in the bistatic plane.

produces the familiar ellipses of constant range sum, or *isorange contours*. A useful relationship is that the bisector of the bistatic angle is orthogonal to the tangent of that ellipse at the target. The tangent is often a good approximation to an isorange contour within the area common to the transmit and receive beams.

When the bistatic radar's receiving antenna is a phased array, and the array normal is also normal to the baseline, θ_R is measured directly by the antenna in any bistatic plane. This fortuitous situation is caused by *conic distortion*, which is inherent in any phased array antenna. However, when the array normal is not normal to the baseline, or when the receiving antenna is mechanically steered or scanned, θ_R is not measured directly. Often DOA measurements are taken in—or converted to—azimuth and elevation angles referenced to an x-y-z coordinate system centered at the receive site, where z is co-linear with the local vertical. Conversion between a north coordinate system and an x-y-z coordinate system is given in Section 5.3 of Willis.[1]

Other coordinate systems, including three-dimensional systems, have been used to define bistatic radar operation.[14,23–29] A polar coordinate system is also shown in Figure 23.1. The (r, θ) coordinates are located on the bistatic plane with origin at the midpoint of the baseline. It is useful for plotting *ovals of Cassini* (see Section. 23.3) and is detailed in Willis.[1] On occasion, the included angles θ_T' and θ_R' are used to define transmitter and receiver look angles in the bistatic triangle, such that $\theta_T' + \theta_R' + \beta = 180°$. In this case, $\theta_T' = 90° - \theta_T$ and $\theta_R' = 90° + \theta_R$ can be used to transform north-referenced equations into included-angle equations. Plots of bistatic clutter data use a separate and quite arcane coordinate system, which is defined in Section 23.8.

Geometry is a principal factor distinguishing bistatic from monostatic radar operation. In evaluating bistatic radar operation, it is useful to start with a geometry-invariant performance measure, which is obtained by setting $L = 0$ (or $R_T = R_R$ and $\beta = 0°$). The result is defined as an *equivalent monostatic range*, or *benchmark range*, and is detailed in Section 23.3. It is also useful as a *sanity check*; because at these limits, all bistatic radar equations must reduce to an equivalent monostatic equation.

23.3 BISTATIC RADAR EQUATION

Benchmark Range Concept. Unlike a monostatic radar, the range performance of a bistatic radar is a function of the geometry, specifically the baseline range L and an antenna look angle, either θ_T or θ_R. When factors such a diffraction, refraction, multipath, and masking are absent or can be ignored, bistatic range as a function of these variables can be plotted on the bistatic plane using an oval of Cassini. An oval of Cassini is the locus of the vertex of a triangle when the product of the sides adjacent to the vertex is constant and the length of the opposite side is fixed. When applied to the bistatic triangle shown in Figure 23.1, the vertex is at the target; R_T and R_R are the sides adjacent to the vertex; and the baseline L is the length of the fixed, opposite side.

Traditionally, ovals of Cassini are drawn as contours of a constant received signal power or a received signal-to-noise ratio around a fixed baseline range, L. Although these signal-dependent contours provide a sense of bistatic radar performance, they do not show maximum/minimum detection ranges and coverage for variable baselines, all parameters of operational interest. To remedy this problem, the concept of a bistatic *benchmark range,* or more simply *benchmark,* is introduced. It is established as follows.[2] First, the bistatic radar range equation is derived in a manner completely

analogous to that for a monostatic radar. The equation is then solved for the *bistatic maximum range product*, $(R_T R_R)_{\max}$. Next, an *equivalent monostatic maximum range,* $(R_M)_{\max}$, is defined omitting the *max* subscript for convenience, as shown here:

$$R_M = (R_T R_R)^{1/2} \tag{23.1}$$

This equivalent monostatic maximum range, also known as the *geometric mean range,* represents performance of the bistatic radar when transmitter and receiver are co-located; i.e., when $L = 0$. It is defined as the bistatic radar's *benchmark range*. Since this benchmark is geometry-invariant, it becomes useful when comparing bistatic to monostatic range performance. Finally, an oval of Cassini is established as a function of the baseline range, L, normalized to the benchmark range, R_M. Based on this oval, maximum and minimum detection ranges and the coverage area are calculated, all as a function of R_M. This procedure is also used to define bistatic operating regions.

Range Equation. The radar range equation for CW or coherent pulse radars[30] is modified for bistatic operation and then solved for the bistatic maximum range product, $(R_T R_R)_{\max}$:

$$(R_T R_R)_{\max} = \left[\frac{P_{\text{av}} t_o G_T G_R \lambda^2 \sigma_B F_T^{\,2} F_R^{\,2}}{(4\pi)^3 k T_o F_n (E/N_o) L_T L_R} \right]^{1/2} \tag{23.2}$$

where

R_T = Transmitter-to-target range (m)
R_R = Receiver-to-target range (m)
P_{av} = Transmitted average power (W)
t_o = Signal observation (or integration) time
G_T = Transmitting antenna power gain
G_R = Receiving antenna power gain
λ = Wavelength (m)
σ_B = Bistatic radar cross section (m^2)
F_T = Pattern propagation factor for transmitter-to-target path
F_R = Pattern propagation factor for receiver-to-target path
k = Boltzmann's constant [1.38×10^{-23} J/K]
T_o = Standard temperature [290 K]
F_n = Receiver noise figure
E/N_o = Received energy to receiver noise spectral density required for detection
L_T = Transmitting system losses (>1)
L_R = Receiving system losses (>1)

Equation 23.2 assumes that a matched filter, or an equivalent matched filter such as a cross-correlator, is used on reception. Equation 23.2 is related to the corresponding monostatic maximum range equation by $R_T R_R = R_M^{\,2}$ and $\sigma_M = \sigma_B$, where σ_M is the monostatic radar cross section. For pulsed radar operation, $t_o = n/f_p$, where n is the number of pulses integrated and f_p is the pulse repetition frequency. Also, the signal-to-noise ratio required for detection, $(S/N)_{\text{req}} = E/N_o$ when $B\tau \approx 1$, where B is the receiver bandwidth and τ is the pulse width.

The signal processing time, t_o, is sometimes set by the amount of doppler spreading or velocity-walk, Δf_d generated by a moving target. Specifically, $\Delta f_d = (t_o)^{-1} = B_n$,

where B_n is the noise bandwidth of receiver's predetection filter. In the monostatic case, doppler spreading is

$$(\Delta f_d)_m = [2a_r/\lambda]^{1/2} \tag{23.3}$$

where a_r is the radial component of target acceleration. Equation 23.3 also applies to the bistatic case at small bistatic angles, β, particularly in an *over-the-shoulder* operation, where the target lies near the baseline extended beyond either the receiver or the transmitter (called the *extended baseline*).* However, for larger β—the general case—the radial component, which is aligned with the bisector of the bistatic angle, will be reduced. A rule of thumb for these large β conditions is

$$(\Delta f_d)_b = [a_r/\lambda]^{1/2} \tag{23.4}$$

Equation 23.4 is used to set t_o and hence the noise bandwidth of receiver's predetection filter B_n. Since $(\Delta f_d)_b > (\Delta f_d)_m$, the constraint on bistatic signal processing time is slightly less than the equivalent monostatic time.

As in the monostatic equation, the transmitting and receiving pattern propagation factors, F_T and F_R, each consist of two terms: the propagation factors, F'_T and F'_R, and the antenna pattern factors, f_T and f_R, respectively. The antenna pattern factors are the relative strength of the free-space field radiated by the transmitting and receiving antennas as a function of their pointing angles. These factors are applied whenever the target is not at the peak of a beam.

Propagation factors customarily include the effects of multipath, diffraction, and refraction, with atmospheric absorption effects included in the loss terms. As with a monostatic radar, bistatic radar propagation requires a suitable path from the transmitter to the target and the target to the receiver. In contrast to a monostatic radar, however, propagation effects can be significantly different over the two bistatic paths and must be treated separately. Multipath is the primary example, where the target can be in a multipath lobe on one path and a multipath null on the other, depending on antenna and target altitude and terrain conditions.

When a correlation receiver uses the demodulated direct path RF signal as its reference, that signal is subjected to interference (multipath and RFI), which is different from interference affecting the target echo path. If the correlator operates in its linear region, the echo plus its interference convolved with the reference plus its interference produces the desired echo with full matched-filter gain plus interference with gain reduced by mismatch. These signals add vectorially to modify the pattern-propagation factor. However, if the correlator operates in its nonlinear region, which can frequently happen, cross-products are generated, which reduce the echo's matched-filter gain. The amount of loss depends on the magnitude of the interference and is accounted for in the signal-processing loss term.[31]

Ovals of Cassini. The free space, maximum detection contour of a bistatic radar's benchmark range is a circle of radius R_M, just as in the monostatic case. Such a circle assumes constant radar cross-section and pattern propagation factors, which are scenario- and geometry-dependent. For the general bistatic case, where $L > 0$, the free-space maximum detection contour becomes the familiar oval of Cassini, again

* In the first case, the transmitter illuminates the target over the receiver's shoulder; in the second case, the receiver views the target over the transmitter's shoulder.

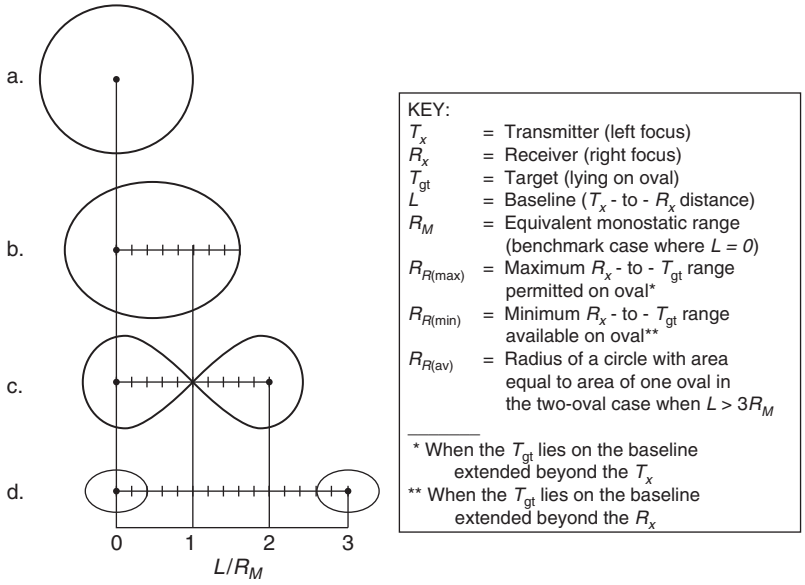

a.
b.
c.
d.
0
1
2
3
L/R_M
KEY:
T_x = Transmitter (left focus)
R_x = Receiver (right focus)
T_{gt} = Target (lying on oval)
L = Baseline (T_x - to - R_x distance)
R_M = Equivalent monostatic range (benchmark case where $L = 0$)
$R_{R(max)}$ = Maximum R_x - to - T_{gt} range permitted on oval*
$R_{R(min)}$ = Minimum R_x - to - T_{gt} range available on oval**
$R_{R(av)}$ = Radius of a circle with area equal to area of one oval in the two-oval case when $L > 3R_M$
* When the T_{gt} lies on the baseline extended beyond the T_x
** When the T_{gt} lies on the baseline extended beyond the R_x

FIGURE 23.2 Normalized ovals of Cassini, lying in the bistatic plane, the plane containing transmitter, receiver, and target[2] (*Courtesy SciTech Inc.*)

with the monostatic caveats cited above. Thus, this oval (or ovals) provides a convenient—but at times over-simplified—view of bistatic range coverage and must be used with care.

An additional bistatic caveat is necessary. When the target is on or near the baseline, i.e., located between receiver and transmitter where the bistatic angle $\beta \rightarrow 180°$, a completely different environment is generated: *forward scatter* from both target and clutter. In this case, the target radar cross section (RCS) and clutter scattering coefficient (σ_o) are greatly enhanced, whereas range and doppler measurements are greatly degraded. Often *normal* bistatic operation excludes this region, so that a 10—20° wedge with the apex at the receiver and directed at the transmitter is excised from the oval. Details are given in Willis.[1]

Figure 23.2 shows four cases of ovals of Cassini normalized to the benchmark range: (*a*) benchmark: $L = 0$; (*b*) one-oval: $L < 2R_M$; (*c*) lemniscate: $L = 2R_M$; and (*d*) two-ovals: $L > 2R_M$. In all cases, the transmitter is located at the left oval focus (0). The receiver is located at the right oval focus (0,1,2,3). Other terms and symbols are defined in the key shown with Figure 23.2.

Table 23.1 lists expressions for calculating oval area and maximum/minimum receiver detection ranges for these four cases, again referenced to the benchmark range, R_M.* For Figure 23.2*d* when $L > 3R_M$, the ovals can conveniently be approximated as circles with radius $R_{R(av)} \sim R_M^2/L$ and corresponding area $\pi R_M^4/L^2$. Expressions for

* The one-oval area formula is derived from Eq. (D.7a) in Willis.[1] It is also used for calculating the lemniscate area. Formulas for the two-oval area are derived from Eq. (D.11a) in Willis.[1] More terms in these series can be used if greater accuracy is required.

TABLE 23.1 Area and Detection Ranges for General Ovals of Cassini[2] (*Courtesy SciTech, Inc.*)

Case	L	Area (of one oval)	$R_{R(\max)}$ (on R_x oval)	$R_{R(\min)}$ (on R_x oval)
Circle (Benchmark)	0	πR_M^2	R_M	R_M
One Oval	$<2R_M$	$\sim\pi [R_M^2 - L^4/ (64\ R_M^2)]$	$(R_M^2 + L^2/4)^{1/2} + L/2$	$(R_M^2 + L^2/4)^{1/2} - L/2$
Two Ovals	$>2R_M$	$\sim\pi R_M^2\ [R_M^2/L^2]$	$L/2 - (L^2/4 - R_M^2)^{1/2}$	$(R_M^2 + L^2/4)^{1/2} - L/2$
	$>3R_M$	$\sim\pi R_M^2\ [R_M^2/L^2]$	$\sim R_M^2/ L$	$\sim R_M^2/ L$

receiver (and transmitter) ranges on the opposite oval are readily calculated through mirror-image symmetry. Also note that the area of every bistatic oval is always less than the monostatic circle.

The expressions in Table 23.1 can also be used to assess first-order bistatic radar line-of-sight (LOS) constraints, where the LOS is defined as a line between transmit and receive antennas tangent to the Earth's surface. Specifically, for a given target, transmitter and receiver altitudes, the target must simultaneously be within LOS to both the transmitter and receiver sites. For a smooth, 4/3 Earth model, the LOS range r_R between a receive antenna of altitude h_R and target of altitude h_t is

$$r_R = 130\ (\sqrt{h_R} + \sqrt{h_t}) \tag{23.5}$$

where all units are in kilometers. Similarly, the LOS range r_T between a transmit antenna of altitude h_T and target of altitude h_t is

$$r_T = 130\ (\sqrt{h_T} + \sqrt{h_t}) \tag{23.6}$$

Thus, to prevent LOS truncation of the ovals, $r_R \geq R_{R(\max)}$ and $r_T \geq R_{T(\max)}$. These expressions ignore both diffraction and multipath, which can significantly alter these ranges, so they must be considered first-order approximations.

A typical task for bistatic air surveillance is to select a baseline L so that a receiver with antenna altitude h_R will match existing LOS coverage of a transmitter with antenna altitude h_T. For the worst case, over-the-shoulder geometry, the requirement would be to match LOS coverage on the extended baseline, $r_T = r_R + L$, so that

$$L = 130\ (\sqrt{h_T} - \sqrt{h_R}) \tag{23.7}$$

For example, when $h_T = 0.1$ km and $h_R = 0.01$ km, $L = 28$ km, which from Eq. 23.5 and 23.6 will provide LOS to an 8.5 m altitude target flying above the extended baseline, 25 km from the receiver and 53 km from the transmitter. Baselines greater than 100 km can pose severe target LOS problems. For example, if $L = 120$ km and $h_T = 0.3$ km, then from Eq. 23.6, the transmitter would only illuminate a target flying directly above the receiver at an altitude $h_t > 0.14$ km. Thus, the target could readily under-fly the illumination, and low-altitude air surveillance capability is lost. As a consequence, the bistatic radar must either employ a greatly elevated (~1 km) transmitter at these long baselines or operate with shorter baselines to achieve acceptable low altitude surveillance coverage. Note that the two-oval case can also require a very high altitude for the site located in the oval not under surveillance—so high that the site must often become airborne.

Finally, the transmit antenna will be in direct LOS of the receive antenna when $L \leq r_T + r_R$ with $h_t = 0$, so that, again with all units in kilometers

$$L \leq 130\ (\sqrt{h_T} + \sqrt{h_R}) \tag{23.8}$$

If Eq. 23.8 is satisfied, extraordinary measures are usually required to suppress the direct path signal to a level where targets can be detected, as outlined in Section 23.9.

23.4 APPLICATIONS

Ovals of Cassini can be used to define three operating regions for a bistatic radar, *co-site*, *receiver-centered*, and *transmitter-centered*.[1] Co-site corresponds to Figure 23.2*b*; receiver-centered to the right oval in Figure 23.2*d*; and transmitter-centered to the left oval in Figure 23.2*d*. The type of transmitter—*dedicated*, *cooperative*, or *non-cooperative*—completes the taxonomy. A dedicated transmitter is designed and controlled by the bistatic or multistatic radar, analogous to a monostatic radar. Both cooperative and noncooperative transmitters are transmitters-of-opportunity—designed for other functions, including radar and communications, but found suitable for bistatic operation. The cooperative transmitter is controlled by allied or friendly forces; the noncooperative transmitter, by hostile or neutral forces. Table 23.2 summarizes bistatic radar applications in these operating regions.*

Entries in the dedicated transmitter/co-site category represent a complete bistatic radar suite—radars with all components including the transmitter designed for bistatic operation. Many of these systems were developed, tested, or deployed prior to 1980. Examples are the French, USSR, and Japanese forward scatter fences used in WWII,[2] the *AN/FPS-23* for air-defense gap filling,[2] *PARADOP* and *MIDOP* range-instrumentation trackers,[9] *SPASUR* for space surveillance,[12,32] and *Sanctuary* for air defense.[33,34] The *BRWL* for artillery, mortar, and rocket location,[35] the Russian *Struna-1* forward scatter fence,[36–38] and the French *Graves* for space surveillance[39] were later developments.

Omitted entries in the dedicated transmitter column for operation in the receiver- and transmitter-centered ovals at large baseline ranges are dictated by operations and cost: both cooperative and noncooperative transmitters-of-opportunity are often present and capable of supporting bistatic operation in these small areas of interest.

TABLE 23.2 Bistatic Radar Applications

Receiver Operating Regions	Range Relationship	Dedicated Transmitter	Cooperative Transmitter	Noncooperative Transmitter
Co-site	$L < 2\,R_M$	• Air surveillance • Range instrumentation • Satellite tracking • Intrusion detection	• Air surveillance • Range instrumentation • Ionospheric measurement • Wind measurement	• Air surveillance
Receiver-centered	$L > 2\,R_M$ $R_T \gg R_R$		• Short-range air surveillance • Silent air-to-ground attack • Planetary exploration	• Short range surveillance
Transmitter-centered	$L > 2\,R_M$ $R_R \gg R_T$		• Planetary exploration	• Air threat monitoring • Missile launch alert

* A multistatic radar with its requirement for common spatial coverage nearly always operates in the co-site region.

Furthermore, this approach is less costly and when exploiting noncooperative transmitters, more covert and less risky than using a dedicated transmitter. Consequently, dedicated transmitters are not usually considered for these applications.

Entries in the cooperative and noncooperative transmitter columns are called *hitchhikers* when the transmitter is a monostatic radar. When the transmitter is from a communications or broadcast system, i.e., not from a radar, entries are called a *passive bistatic radar (PBR).*

Examples of a hitchhiker operating with a cooperative transmitter in the co-site region are the *Multistatic Measurement System* operating with *TRADEX* for increasing ballistic missile reentry measurement accuracy[40,41] and the commercial, Binet Inc., bistatic receiver operating with monostatic doppler weather radars to measure three-dimensional vector wind fields.[2,42,43] Figure 23.3 is a block diagram of the Binet Inc.-developed prototype.

Examples of hitchhikers operating with a cooperative transmitter in the receiver-centered region are the *Covin Rest* program, operating with the space-shuttle radar for synthetic aperture radar (SAR) ground mapping[44]; the *TBIRD* program operating with *Joint STARS* for silent air-to-ground attack via forward-looking bistatic SAR[2,45]; and the *BAC* program operating with *AWACS* for alerting and cueing short range, mobile air-defense systems.[46,47]

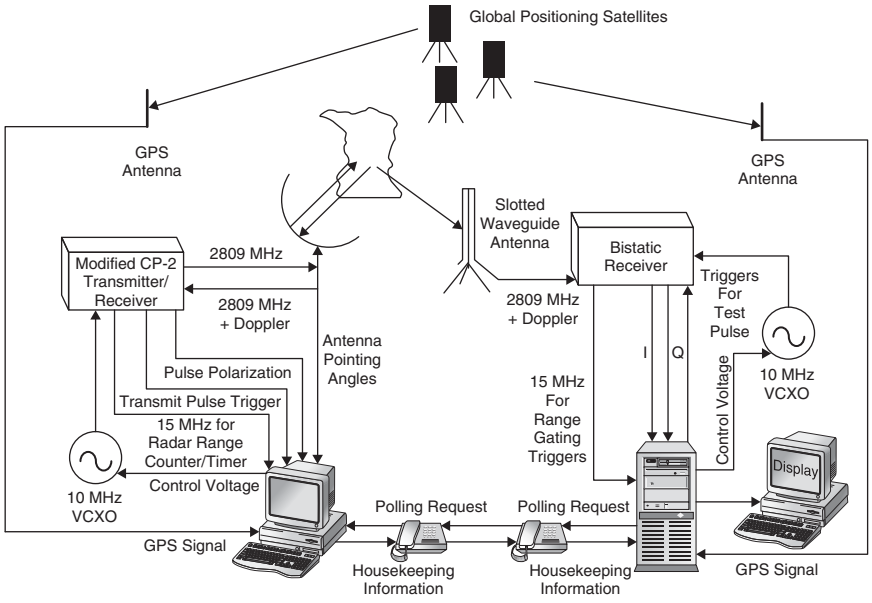


FIGURE 23.3 Simplified block diagram of the Binet Inc.-developed bistatic wind measurement prototype system. A modified CP-2 monostatic weather radar transmits and receives on a narrow beam antenna, while one or more bistatic receive sites 10–20 km from the CP-2 receive bistatically scattered energy from the same illuminated weather volume over a broad antenna beam. A transmit/receive tube is used to protect the receiver, just as in a monostatic radar. All generated frequencies are locked to the master 10-MHz VCXO, which is, in turn, locked to GPS timing signals. Synchronization and other housekeeping data are sent over telephone lines. Signal and data processing is PC-based. (*after J. Wurman[43] © IEEE 1994*)

Hitchhikers using cooperative monostatic radar transmitters also have the inherent capability to counter retro-directive jammers operating against their host radar. Because the jammer uses a high-gain antenna to retro-direct the transmitter's signal back to the transmitter (and thus its monostatic receiver), the spatially separated hitchhiker can be positioned in the sidelobes of that antenna, thus reducing the effectiveness of the jammer. A rule of thumb is to anticipate enhanced hitchhiker performance whenever the bistatic angle is greater than the estimated 3-dB width of the retro-directed main beam.

Examples of PBRs operating with cooperative broadcast transmitters in the co-site region are the *Manastash Ridge Radar* operating with an FM broadcast transmitter for tropospheric soundings,[2,48] *Silent Sentry* operating with FM and TV broadcast transmitters,[49,50] and the *HDTV-Based Passive Radar* operating with a high definition-TV broadcast transmitter,[51] the latter two configured for air surveillance. Examples of PBRs operating with cooperative communications transmitters in the receiver- and transmitter-centered regions are bistatic radars for planetary exploration. They use a data link transmitter on the probe vehicle in the transmitter-centered region and an Earth-based command transmitter in the receiver-centered region.[2,52]

The term cooperative transmitter is somewhat of a misnomer. For example, a cooperative TV or FM station operator would not be inclined to "cooperate" with a PBR by altering antenna coverage or modifying broadcast material with special waveforms. Furthermore, there is always the possibility of a transmitter failure in the course of normal operations or from an enemy attack. This event will degrade performance of a PBR that operates with multiple transmitters and eliminate performance of a PBR that operates only with that transmitter. As a result, the PBR is free to exploit cooperative (or noncooperative) transmissions if they are suitable; however, commerce controls their operation with the PBR remaining a user of opportunity, specifically suffering nonoptimum waveforms (Section 23.9), limited elevation coverage, and occasionally reduced or denied performance.

Constraints also apply when a hitchhiker attempts to exploit a cooperative or noncooperative monostatic radar, specifically with the hitchhiker suffering antenna scan-on-scan problems (Section 23.9). However, the potential for enhancing performance of a cooperative monostatic radar against retro-directive-type jammers when using a suitably positioned hitchhiker necessitates distinguishing cooperative from noncooperative transmitters.

In the noncooperative transmitter column, if a noncooperative transmitter and a hitchhiker were located in or near a battle area, a hitchhiker could use that transmitter in the co-site region just as it would for a cooperative transmitter. The German *Klein Heidelberg* hitchhiking off the British *Chain Home* radars to conduct air surveillance during WWII is an example.[53,54] A hitchhiker implanted in or flying over hostile areas could use any high-powered satellite transmitter illuminating that area to conduct short-range surveillance in the receiver-centered oval.

The most significant system to use a dedicated transmitter is the *Space Surveillance (SPASUR)* 217 MHz multistatic radar fence. It was deployed starting in 1958 at seven sites spanning the continental United States to detect and track noncooperative satellites.[12,32] Transmitters are located at three sites, the largest of which transmits 1 MW CW from a linear array about 3 km long, generating a fixed fan beam. The six receive sites consist of seven or eight linear arrays with dimensions on the order of 1 km, also generating fixed fan beams collinear with the transmit beam. Figure 23.4 shows the data flow in a typical receiving station.[12]

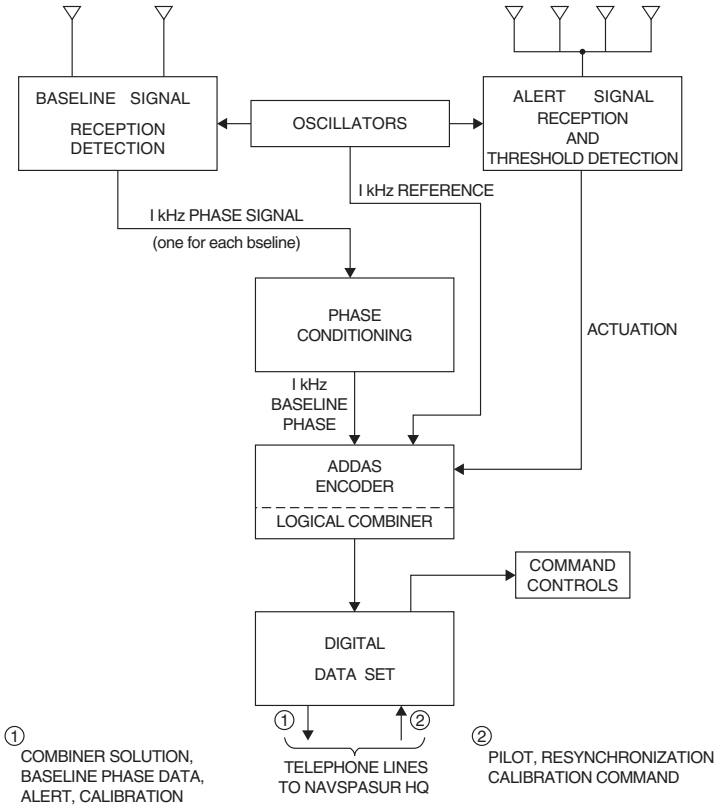


FIGURE 23.4. Real-time data flow in a NAVSPASUR receiving station. " Transmitter energy reflected by the satellite is received by various collinear arrays of dipoles at the receiver station… [Signals] from four in-line arrays feeds the alert receiver which detects the presence of radio energy in excess of preset threshold levels. The reflected energy is also simultaneously received on other arrays that are separated in the east-west direction by various distances (baseline). Signals from any two can be added to form a baseline pair. The baseline receivers employ triple frequency conversion where the phase difference of the signals received by the antenna pair is preserved in a 1000-hertz difference frequency [called a 1 kHz phase signal in the figure]… There is a radio receiver and consequently a 1000-hertz phase [signal] for each baseline pair (12 east west and 3 north south). This phase signal is compared with a 1000-hertz reference signal and encoded (digitized) by the ADDAS encoding equipment. The digital combiner, contained within the ADDAS encoder, receives the 1 KHz phase data and generates an unambiguous zenith angle solution. The combiner outputs along with digitized phase information and control bits, foremost of which is 'alert,' are applied to the telephone lines via a digital data transmitter."(*from* Handbook for NAVSPASUR System Orientation,[12] *courtesy U.S. Navy*)

Target location is established by triangulation, i.e., the intersection of zenith angle (DOA) measurements from two or more receive sites. Subsequently a three-site fence in South Texas was deployed to evaluate the use of bistatic range measurements to improve location accuracy,[55] but it never entered continuous operation.

According to Easton,[32] the *SPASUR* design was driven by cost: a 15,000 nmi detection range required very high average power, which was satisfied by the lowest cost CW operation. But this solution, in turn, required separate sites for isolation—hence, multistatic operation. The stationary beam linear arrays also minimized cost when compared to scanning arrays or reflector antennas. In short, fix the beams and let the satellites fly through them. It has been in continuous use since 1958.

A hitchhiker using a cooperative radar transmitter in the co-site region was the *Multistatic Measurement System (MMS)*. It was installed at the U.S. Kwajalein Missile Range in 1980 as an adjunct to the *TRADEX* L band monostatic radar.[40] *TRADEX* operated in its normal monostatic mode, acquiring, tracking, and illuminating ballistic missile reentry vehicles (RVs). Two unmanned, slaved receive stations, located about 40 km from *TRADEX*, received echoes bistatically scattered off the RV and recorded bistatic range, doppler, and signature data from RVs. This data was used to calculate RV position and dynamics near the atmospheric pierce point or start of re-entry. The system was projected to measure three-dimensional position and velocity with accuracies better than 4 m and 0.1 m/s, respectively, throughout re-entry.[43]

Field tests showed that MMS range data combined with *TRADEX*'s monostatic range data in a trilateration net provided the most accurate estimate of exo-atmospheric RV positions obtained by any of the range radars. MMS operations were concluded in 1993[56] after the metric accuracy of monostatic radars was improved.[57]

In the mid-1990s, a PBR using a cooperative FM broadcast transmitter in the co-site region was developed by the University of Washington. Called the *Manastash Ridge Radar (MRR)*, it is designed to study turbulence in the ionosphere, specifically auroral E-region irregularities, using range, doppler, and DOA (via interferometry) measurements.[2,48] Motivation for MRR development included lower cost, increased safety, spectrum availability, and pedagogical opportunity. MRR provides range-time intensity and range-doppler plots to the World Wide Web every half-hour. Although it is not subject to the stringent air-defense requirements for accurate location of multiple targets in real time, it has detected meteors and aircraft in the course of normal operations.

A second PBR, called the *HDTV-Based Passive Radar*,[51] exploits a cooperative high-definition TV transmitter for air surveillance in the co-site region. It uses range multilateration from four receivers located within 10 km of the transmitter to track low-flying aircraft and helicopters as a gap-filler for monostatic air surveillance radars. Predicted detection and tracking ranges of 30 km on a 1 m^2 target have been demonstrated in real time with 2D tracking errors generally less than 50 m. Coarse target elevation has also been measured. Doppler data has been used to resolve *ghosts*, i.e., false detections that inevitably arise when using multilateration on unaugmented targets.[2]

PBRs have used satellite communication transmitters to measure characteristics of moon and planetary surfaces and atmospheres in both the transmitter- and receiver-centered regions. The first successful piggyback operation in 1967 exploited data link signals transmitted from the *Luna-11* probe, scattered off the moon's surface, and then received by an Earth-based station, in a down-link mode characterized by the transmitter-centered oval of Cassini. Such stations include the Arecibo Observatory and the NASA Deep Space Network. Subsequent measurements were made using *Lunar Orbiter-1*, *Explorer-35*, and *Apollo 14–16*. Mars bistatic radar measurements were made using *Mariner-6*, *-7*, *Viking-1*, *-2*, *Mars Global Surveyor*, and *Mars Express*. Venus measurements were made by *Veneras-9*, *-10*, *Magellan*, and *Venus Express*.[2,52] Figure 23.5 is a simplified schematic of an Earth-based receiver.[2]

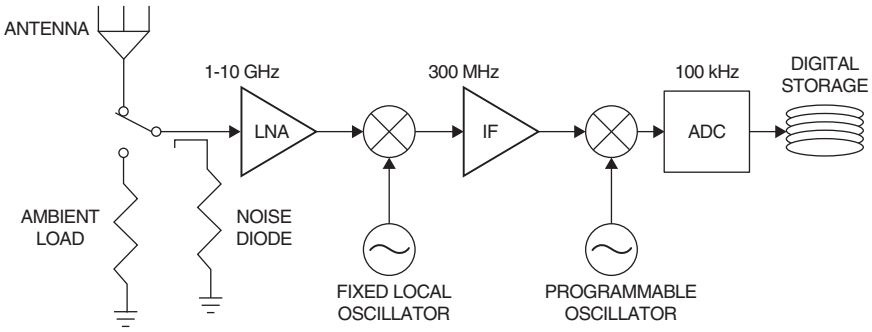


FIGURE 23.5. Block diagram of a typical Earth-based system used for downlink bistatic radar. The low-noise amplifier (LNA) is either a cooled maser or a field effect transistor and may be switched between the antenna and an ambient load, allowing amplitude calibration of the input. During bistatic operations, a signal from the low-level noise diode, previously calibrated against the ambient load, may be injected to monitor real-time performance of the system. Microwave inputs (1–10 GHz) are mixed to a 300 MHz intermediate frequency (IF) for amplification. A programmable oscillator, which can correct for first-order doppler effects, then mixes the IF signal to baseband, and digital samples are stored for later processing. Although an analog-to-digital conversion is shown at the output, the ADC may actually take place at any point in the system.[2] (*Courtesy SciTech*)

A reciprocal but more complex and costly uplink mode uses a bistatic receiver carried by the probe, collecting high power, Earth-transmitted command signals first scattered off the planet's surface and characterized by the receiver-centered oval of Cassini. It has an ~30 dB greater link margin, which was first used for Mars Odyssey and is planned for future probes. In both configurations, even though two legs of the bistatic triangle are extraordinarily long (>10^5 miles), the third leg is sufficiently short (~10 miles) to produce strong echoes at the receiver.[2]

These piggyback bistatic radars have provided useful data in simple, inexpensive surveys of planetary surface properties as a prelude to robotic or human exploration, specifically centimeter- to meter-scale roughness and material densities in the top few centimeters of regolith. Special geometries, such as near-backscatter to identify deposits of clean water ice, are also a unique advantage of bistatic radar. The ability to observe forward scattering at latitudes away from planetary equators is also advantageous for probing surface characteristics.

23.5 BISTATIC DOPPLER

Figure 23.6 defines the geometry for bistatic doppler when the target, transmitter, and receiver are moving. The target has a velocity vector of magnitude V and aspect angle δ referenced to the bistatic bisector. The transmitter and receiver have velocity vectors of magnitude V_T and V_R and aspect angles δ_T and δ_R, respectively, referenced to the north coordinate system of Figure 23.1. All vectors are projections of the three-dimension vectors onto the bistatic plane.

Target Doppler. When the transmitter and receiver are stationary ($V_T = V_R = 0$), the target's bistatic doppler at the receive site f_B is

$$f_B = (2V/\lambda) \cos \delta \cos (\beta/2) \tag{23.9}$$

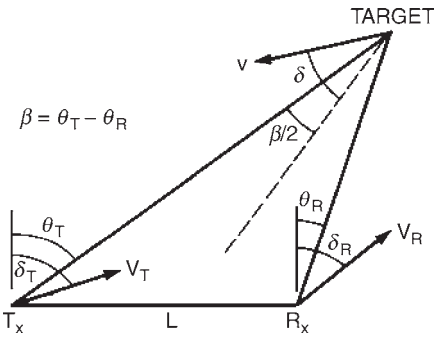


FIGURE 23.6 Geometry for bistatic doppler in the bistatic plane

The term f_B also defines the *doppler beat frequency*, which is produced by mixing the target's doppler with the direct path signal in the receiver. Willis[1] provides an expression for f_B when all three sites are in motion. Equation 23.9 shows that

- When $\beta = 0°$, f_B reduces to the monostatic case for a monostatic radar located on the bistatic bisector. The magnitude of the bistatic doppler is never greater than that of this monostatic doppler.
- When $\beta = 180°$, $f_B = 0$ for any δ, which is the forward-scatter case.
- When $\delta = \pm 90°$ the bistatic doppler is zero. Since these velocity vectors are also tangent to a range-sum ellipse at this point, all such ellipses (including the baseline) become contours of zero target doppler.
- When $\delta = 0°$, the bistatic doppler is a maximum. Since this velocity vector is also tangent to a hyperbola orthogonal to the range-sum ellipse at this point, all such hyperbolas become contours of maximum target doppler.
- When $\delta = \pm \beta/2°$, the velocity vector is pointed at the transmitter or receiver and $f_B = (2V/\lambda)\cos^2(\beta/2)$, which occasionally appears in the literature as a special case of Eq. 23.9.

If a monostatic radar is located at the transmit site and a bistatic hitchhiker is located at the receive site, both measuring target doppler, f_M and f_B, respectively, the two measurements can be combined to estimate the target velocity vector (V, δ) in the bistatic plane. One such estimate is

$$\delta = \tan^{-1}\{[f_M / f_B \sin(\beta/2)] - \cot(\beta/2)\} \tag{23.10}$$

$$V = \lambda f_B / 2\cos\delta = \lambda f_M / 2\cos(\delta - \beta/2) \tag{23.11}$$

where β is obtained by solving the bistatic triangle, for example, by using monostatic range, monostatic LOS, and an estimate of the baseline. A third hitchhiking site allows the target velocity vector to be measured in three dimensions. This process is called *dual-doppler* for two measurements and *multiple-doppler* for three or more measurements, and has been used to measure three-dimensional vector wind fields.[2,42,43,58]

Isodoppler Contours. When the target is stationary and the transmitter and receiver are moving (e.g., on airborne platforms), the bistatic doppler shift at the receiver site f_{TR} is

$$f_{TR} = (V_T/\lambda) \cos (\delta_T - \theta_T) + (V_R/\lambda) \cos (\delta_R - \theta_R) \tag{23.12}$$

where terms are defined in Figure 23.6.

The locus of points for constant doppler shift on the Earth's surface is called an *isodoppler contour*, or *isodop*. Clutter returns are characterized by these isodops, which are called *clutter doppler shift*. In the monostatic case and a flat Earth, these isodops are conic sections in three dimensions and radial lines emanating from the radar in two dimensions. Because these isodops are aligned with the radar's look angle, the clutter is called *stationary*. In the bistatic case, the isodops are skewed away from the look angle, depending upon the geometry and platform motion, and the clutter is called *nonstationary*. Bistatic isodops are developed analytically for two dimensions and a flat earth by setting f_{TR} = constant in Eq. 23.12 and solving for θ_R (or θ_T, if appropriate).

Figure 23.7 is a plot of bistatic isodops in a two-dimensional bistatic plane, i.e., where the transmitter and receiver are at zero or near-zero altitude, for the following conditions: $V_T = V_R = 250$ m/s, $\delta_T = 0°$, $\delta_R = 45°$, and $\lambda = 0.03$ m.

Dimension of the grid on the bistatic plane is arbitrary; that is, the isodops are invariant with scale. On the left and right sides of Figure 23.7, the isodops are approximately stationary, which are pseudo-monostatic operating points. Elsewhere, the isodops are nonstationary. In these nonstationary regions, the quality of bistatic SAR imagery is limited and moving target indication (MTI) performance is degraded when using standard monostatic radar processing techniques.

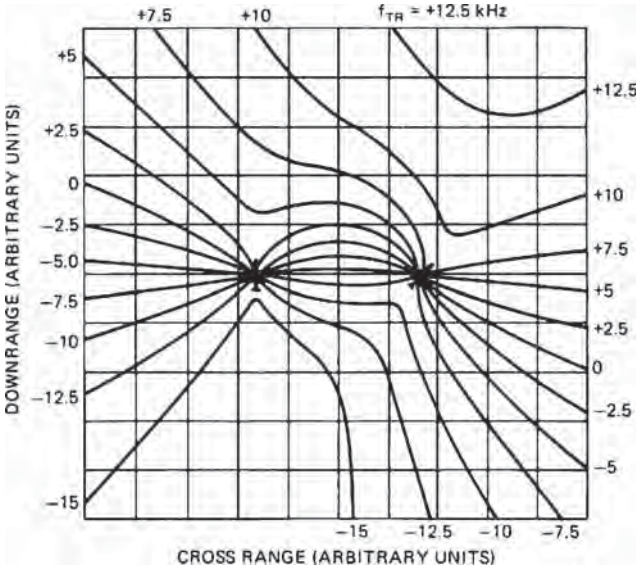


FIGURE 23.7 Bistatic isodoppler contours for two dimensions and a flat earth (*Courtesy Lee R. Moyer, Technology Service Corp.*)

Research was started in the 1990s to improve the quality of bistatic SAR images, which were constrained to coherent integration times of a few seconds[1] and consequently were of limited tactical interest. Specifically, local oscillator phase instabilities were reduced and bistatic automatic focusing (*autofocus*) algorithms were developed to improve range measurement accuracy from the antenna phase center to the imaged scene. Bistatic autofocus requires that the position of both transmit and receive platforms must be tracked with subwavelength relative accuracy to correct time-varying phase errors as part of the image formation process. The result allowed an increase in coherent integration times comparable to monostatic SAR, typically >10 seconds. Consequently, bistatic SAR image quality was greatly improved, as reported by B. Rigling in Chapter 10 of Willis and Griffiths.[2]

Bistatic space-time adaptive processing (STAP) was also developed to improve MTI performance of separately moving transmit and receive platforms. Since bistatic clutter exhibits nonstationary space-time characteristics, bistatic STAP methods are not simple applications of monostatic approaches, but a new class of algorithms. Specifically, they apply a data-dependent weighting to voltages collected over multiple receive antenna channels and pulses. This weighting dynamically tailors a filter response in angle and bistatic doppler to suppress ground clutter reflections. Digital beamforming[59,60] is required for this operation. Other necessary elements include a means to estimate the spatio-temporal clutter covariance matrix (the data-dependent element of the filter weighting) and hypothesis of the target steering vector. In general, auxiliary data taken from range bins other than the cell under test is used to estimate the unknown, but critical, clutter covariance matrix. With compensation for nonstationary behavior, bistatic clutter suppression has been greatly improved, as reported by W. Melvin in Chapter 11 of Willis and Griffiths.[2]

23.6 TARGET LOCATION

Bistatic Location. A bistatic receiver typically uses range sum $(R_T + R_R)$ for target location, which can be estimated by two methods. In the direct method, the receiver measures the time interval ΔT_{rt} between reception of the transmitted signal and reception of the target echo. It then calculates the range sum as $(R_T + R_R) = c\Delta T_{rt} + L$. This method can be used with any suitable modulated transmission and any type of transmitter (dedicated, cooperative, or noncooperative), given an adequate LOS between transmitter and receiver. In the indirect method, presynchronized stable clocks are used by the receiver and a dedicated transmitter. The receiver measures the time interval ΔT_{tt} between the transmission of a signal and reception of the target echo. It then calculates the range sum as $(R_T + R_R) = c\Delta T_{tt}$. A transmitter-to-receiver LOS is not required unless periodic clock synchronization is implemented over the direct path.

The traditional method for converting target range sum into target range from the receiver, R_R, is[13]

$$R_R = \frac{(R_T + R_R)^2 - L^2}{2(R_T + R_R + L\sin\theta_R)} \tag{23.13}$$

The baseline L can be determined using GPS or other methods such as an emitter locator for noncooperative transmitters. As outlined in Section 23.2, the receiver look angle, θ_R, can be measured directly with a phased array antenna that scans in two dimensions,

or target azimuth and elevation measurements can be converted to θ_R.[1] Beam-splitting techniques can be used to reduce the measurement error. Even with beam-splitting, θ_R is the critical parameter establishing accuracy of the R_R estimate, since, as in the monostatic case, its error is proportional to target range. A full error analysis of Eq. 23.13 is given in Section 5.2 of Willis.[1] No bistatic radar operating autonomously has been shown to provide adequate location of air or space targets without employing receive apertures comparable in size to monostatic receive apertures used for those purposes.

In the special case of a bistatic radar using the direct range sum estimation method, when $R_T + R_R \approx L$ Eq. 23.13 can be approximated as[1]

$$R_R = \frac{c\Delta T_{\text{rt}}}{1 + \sin\ \theta_R} \tag{23.14}$$

Two examples are a receiver operating in an over-the-shoulder geometry and the transmitter operating on a satellite with the receiver and target near the earth. The error in Eq. 23.14 is less than 10 percent for $0° < \theta_R < 90°$ and $L > 0.82\ (R_T + R_R)$ or $4.6\ c\Delta T_{\text{rt}}$. The error grows rapidly for $\theta_R < 0°$.

Other target location techniques are possible.[13,14,18,29,61–64] For example, when a bistatic hitchhiker exploits the transmitter of a monostatic radar, the radar's look angle θ_T can be used in place of or in conjunction with θ_R. An example of the latter is the *theta-theta* location technique, where

$$R_R = L \cos\theta_T / \sin\ (\theta_T - \theta_R) \tag{23.15}$$

and $\theta_T - \theta_R = \beta$. A dedicated or cooperative monostatic radar can provide values of θ_T directly to the hitchhiker. Otherwise, the hitchhiker must estimate the value, for example, via an emitter locator measuring the radar's antenna scan rate when it is predictable. In this case, target location accuracy is often set by the θ_T estimation error.

Multistatic Location.[9,14,65] Multistatic location typically uses multiple transmitters operating with one receiver or multiple receivers operating with one transmitter. Ellipses of constant range-sum, i.e., isorange contours, from each transmitter-receiver pair are calculated and combined at a central site to produce intersecting contours, which locate the target.* Multistatic cross-range location can be more accurate than monostatic or bistatic cross-range location since angle data, with its range-dependent accuracy, is not used. The multistatic radar must, however, use multiple, properly located sites with both overlapping coverage and simultaneous measurements, which, in turn, require broad transmit and receive beams to achieve this accuracy. These requirements usually combine to restrict multistatic air surveillance performance to short or medium ranges.

Geometric dilution of precision (*GDOP*) establishes multistatic location accuracy (and resolution) and is developed by D. Barton in Chapter 6 of Willis and Griffiths.[2] GDOP is a function of the angle of intersection, α, between isorange contours. Because the bisector of a bistatic angle is orthogonal to an isorange contour, GDOP can readily be determined by the angle of intersection, also α, of these bistatic bisectors. In the simplest case, the down-range error σ_{dr} is proportional to $[\sqrt{2} \cos\ (\alpha/2)]^{-1}$ and cross-range error σ_{cr} is proportional to $[\sqrt{2} \sin\ (\alpha/2)]^{-1}$.

* Isorange contours intersect at other locations as well. These nontarget locations are called *ghosts,* which must be excised. This subject is treated in Chapter 6 of Willis and Griffiths.[2]

For example, when the target is surrounded on three sides by a receive site, a transmit site, and another receive site such that $\alpha = 90°$, $\sigma_{dr} = \sigma_{cr} = 1$. This geometry represents the optimum case of a unity GDOP factor, yielding a circular error ellipse with radius equal to the range error of one transmit-receive pair. In contrast, when the target is located some distance from the three sites, α is reduced. For example, when $\alpha = 5°$, $\sigma_{dr} = 0.71$, and $\sigma_{cr} = 16.2$. Thus, the down-range error is slightly reduced but the cross-range error is greatly increased, not unlike that of a radar using angle data to establish cross-range accuracy.

These examples also apply when ground-based multistatic sites attempt to measure target altitude. When the sites surround the target, for example, when positioned around the launch site of a ballistic missile, α remains relatively large during the missile's launch phase, yielding precise altitude estimates. When the sites are some distance from the target, for example, when conducting air or missile surveillance, α is small, yielding poor altitude estimates.

Both bistatic and multistatic radars can potentially achieve even better location accuracy by using narrow-band doppler tracking under the following conditions: (1) When integrating doppler data, initial conditions (the pesky constant of integration) can be established with sufficient accuracy. (2) When taking sequential doppler measurements, the target's velocity vector remains constant[64] or predictable.[62] For example, many doppler-only, precision range instrumentation systems were developed in the U.S. after WWII.[9] Both beacon-aided and skin-track systems were developed. All required initialization of the track data, which was conveniently provided by the target's launch coordinates. However, if the target return momentarily faded or the transponder signal was interrupted during flight so that track was lost, there was no way to reinitialize new track data and subsequent location estimates became biased or were lost. These systems were subsequently replaced with precision monostatic radars and optical trackers.

23.7 TARGET CROSS SECTION[14,15,30,66–87]

The bistatic radar cross section (RCS) of a target σ_B is a measure, as is the monostatic radar cross section σ_M, of the energy scattered from the target in the direction of the receiver. Bistatic cross sections are more complex than monostatic cross sections in the optical region since σ_B is a function of aspect angle and bistatic angle β.* Three bistatic RCS regions are of interest in the optical region: pseudo-monostatic, bistatic, and forward scatter. Each region is defined by the bistatic angle. The extent of each region is set primarily by the target's physical characteristics.

Pseudo-Monostatic RCS Region. The Crispin and Siegal monostatic-bistatic equivalence theorem applies in the pseudo-monostatic region[69]: for vanishingly small wavelengths, the bistatic RCS of a sufficiently smooth, perfectly conducting target is equal to the monostatic RCS measured on the bisector of the bistatic angle. Sufficiently smooth targets include spheres, elliptic cylinders, cones, and ogives, allowing the region to extend out to $\beta = 40°$ and occasionally out to $\beta = 90°$.[1,76–79]

* However in the resonance region, typically at VHF and low UHF for many air targets, variations in $\beta < \sim 90°$ have little effect on σ_B, such that $\sigma_B \approx \sigma_M$.[2]

For targets of more complex structure, the extent of the pseudo-monostatic region is reduced. A variation of the equivalence theorem developed by Kell[74] applies to small bistatic angles, in some cases as small as 5°: the bistatic RCS of a complex target is equal to the monostatic RCS measured on the bisector of the bistatic angle at a frequency lower by a factor of cos (β/2).

Kell's complex targets are defined as an assembly of discrete scattering centers (simple centers such as flat plates, reflex centers such as corner reflectors, skewed reflex centers such as a dihedral with corner ≠ 90°, and stationary phase regions for creeping waves). When the wavelength is small compared to target dimensions, these complex target models approximate conventional aircraft, ships, ground vehicles, and some missiles. The targets can be composed of conducting and dielectric materials.

At small bistatic angles, the cos (β/2) frequency reduction term has little effect in Kell's pseudo-monostatic region. For example, a 10° bistatic angle corresponds to a 0.4% shift in wavelength and usually can be ignored. Both versions of the equivalence theorem are valid when the positions of the transmitter and receiver are interchanged, given that the target-scattering media are reciprocal. Most media are reciprocal. Exceptions are gyrotropic media, such as ferrite materials and the ionosphere.

Bistatic RCS Region. The bistatic angle at which the equivalence theorem fails to predict the bistatic RCS identifies the start of the second bistatic region. In this region, the bistatic RCS diverges from the monostatic RCS. Kell[74] identified three sources of this divergence for complex targets and for a target aspect angle fixed with respect to the bistatic bisector. These sources are (1) changes in relative phase between discrete scattering centers, (2) changes in radiation from discrete scattering centers, and (3) changes in the existence of centers—appearance of new centers or disappearance of those previously present.

The first source is analogous to fluctuations in monostatic RCS as the target aspect angle changes, but now the effect is caused by a change in bistatic angle.[87] The second source occurs when, for example, the discrete scattering center, including flat plates, retro-reflects energy toward the transmitter, and the receiver is positioned outside the retro-reflected beamwidth; thus, the received energy is reduced. The third source is typically caused by shadowing, for example, by an aircraft fuselage blocking one of the bistatic paths—transmitter or receiver LOS to a scattering center.

In general, this divergence results in a bistatic RCS lower than the monostatic RCS for complex targets. For example, Ewell and Zehner[81] measured the monostatic and bistatic RCS of coastal freighters at X band when both the transmitter and the receiver were near grazing incidence. The data was plotted as a ratio of bistatic to monostatic RCS, σ_B/σ_M. The measurements generally match Kell's model: of the 27 data points, 24 show bistatic RCS lower than monostatic RCS. The reduction in bistatic RCS starts between $\beta = 5°$ and $\beta = 10°$ and trends downward to $\sigma_B/\sigma_M = -15$ dB at $\beta = 50°$.

This rather severe loss was measured under special conditions: low grazing angles for targets with vertical surfaces, and dihedrals and trihedrals, which generate a large monostatic RCS. Thus, the bistatic RCS becomes significantly lower as the bistatic angle increases due to shadowing and loss of these speculars and retro-reflectors. Bistatic losses should not be as severe for targets with blended surfaces and a less complex structure, such as combat aircraft.

Glint Reduction in the Bistatic RCS Region. A second effect can occur in the bistatic region. When the bistatic RCS reduction is caused by a loss or attenuation of

large discrete scattering centers, for example, through shadowing, target glint is often reduced. Target glint is the angular displacement in apparent phase center of a target return and is caused by the phase interference between two or more dominant scatters within a radar resolution cell. As the target aspect angle changes, this phase interference changes, shifting the apparent phase center, often with excursions beyond the physical extent of the target. These excursions can significantly increase the errors in angle tracking or measurement systems. When the returns from dominant scatterers are reduced in the bistatic region, the source and hence the magnitude of glint excursions is reduced. Measurements with tactical aircraft showed that for a 30° bistatic angle, peak glint excursions could be reduced by a factor of 2 or more, with most of the excursions contained within the physical extent of the target.[88] This reduction can be exploited in a semi-active homing missile by modifying its trajectory to maintain $\beta > 20$–$30°$ during endgame.

Forward-Scatter RCS Region. The third bistatic RCS region, forward scatter, occurs when the bistatic angle approaches 180°. When $\beta = 180°$ Siegel[66] showed, based on physical optics, that the forward-scatter RCS, σ_F, of a target with silhouette (or shadow) area A is

$$\sigma_F = 4\pi A^2/\lambda^2 \tag{23.16}$$

where λ, the wavelength, is small compared with the target dimensions. The targets can be either smooth or complex structures and, from the application of Babinet's principle, can be totally absorbing.[70,75]

For $\beta < 180°$, the forward-scatter RCS rolls off from σ_F. The rolloff is approximated by treating the shadow area A as a uniformly illuminated antenna aperture. The radiation pattern of this *shadow aperture* is equal to the forward-scatter RCS rolloff when $(\pi - \beta)$ is substituted for the angle off the aperture normal. A sphere of radius a will roll off 3 dB at $(\pi - \beta) \approx \lambda/\pi a$, when $a/\lambda \gg 1$. The roll-off continues approximating $J_0(x)/x$ down to $\beta \approx 130°$, where J_0 is a Bessel function of zero order. A linear aperture of length D, with aspect angle orthogonal to the transmitter LOS, will roll off 3 dB at $(\pi - \beta) = \lambda/2D$, when $D/\lambda \gg 1$. The forward-scatter RCS rolloff continues, with sidelobes approximating $\sin x/x$ over the forward-scatter quadrant ($\beta > 90°$).[30] For other aspect angles and targets with complex shadow apertures, calculation of the forward-scatter RCS rolloff usually requires simulation.

The forward-scatter RCS of more complex bodies has been simulated and measured; the bodies were both reflecting and absorbing.[67,70,71,76,82,84–86] Figure 23.8 shows a method-of-moments simulation of a 16- by 1.85-cm cylinder with 992 facets at 35 GHz, for three fixed transmitter-to-target geometries: (*a*) near end on, (*b*) 45° aspect angle, and (*c*) broadside.[84] All three bistatic regions are shown in the figure. In the broadside geometry, the pseudo-monostatic region occurs at $\beta < 20°$, bistatic at $20° < \beta < 140°$, and forward scatter at $\beta > 140°$. The other two geometries show a similar but broader forward-scatter lobe, as is expected since the silhouette area and hence the shadowing aperture are smaller. The 45° aspect geometry is of interest because the RCS in the bistatic regions is larger than the monostatic RCS for most bistatic angles. The large spike at $\beta = 90°$ is the bistatic specular lobe, analogous to the monostatic specular lobe in the broadside geometry. While Figure 23.8 shows the clear dependency of bistatic RCS on both aspect and bistatic angles, it also serves to caution against attempts to use oversimplified bistatic RCS models, especially in the bistatic region.

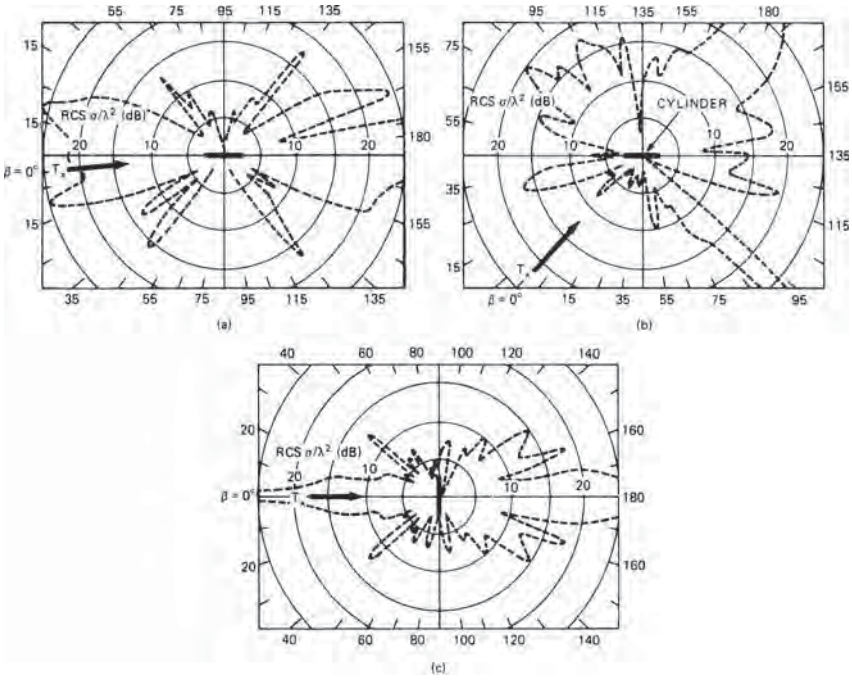


FIGURE 23.8 Simulated bistatic RCS, replotted as a function of bistatic angle for a conducting cylinder, 16 by 1.85 cm at 35 GHz, HH polarization (*after R. C. Paddison et al.*[84]): (*a*) near end on, (*b*) 45° aspect angle, and (*c*) broadside

23.8 SURFACE CLUTTER

The bistatic radar cross section of surface clutter, σ_c, is a measure, as is the monostatic radar clutter cross section, of the energy scattered from a clutter cell area, A_c, in the direction of the receiver. It is defined as $\sigma_c = \sigma_B^0 A_c$, where σ_B^0 is the scattering coefficient, or the clutter cross section per unit area of the illuminated surface. The clutter cell area A_c has been developed for beam- and range-limited cases and is reported elsewhere.[1,2] Doppler-limited cases are a function of platform motion, which, in turn, depend on a specific scenario. Consequently, they are modeled on a case-by-case basis.

Bistatic Scattering Coefficient. Values of the scattering coefficient σ_B^0 vary as a function of the surface composition, frequency, and geometry and are obtained through field measurement programs. In 1981, M. M. Weiner[89] documented and evaluated all unclassified measurements of σ_B^0; however, its use was restricted to U.S. government agencies. In 1990, Willis used Weiner's references to reconstruct and evaluate typical data from Weiner's work,[89] which became available for public use in Willis.[1,16] In 2003, Weiner's work[89] was cleared for public release, making available in one document all unclassified σ_B^0 data and analysis through 1980. Weiner then updated his work with data available through 2005 and republished it in Chapter 9 of Advances in Bistatic Radar.[2] This section summarizes and comments on essential elements of Weiner's work.

The available database for terrain and sea clutter at microwave frequencies consists of nine measurement programs, which are summarized in Table 23.4. The measurement angles shown in Table 23.3 are defined in Figure 23.9, which is a clutter-centered coordinate system similar to those used in all the measurement programs. Because terrain and sea are reciprocal media, θ_i and θ_s are interchangeable in the subsequent data.

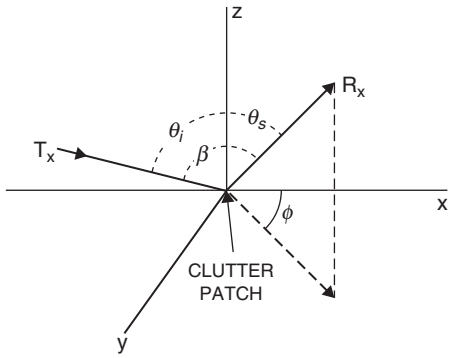


FIGURE 23.9 Coordinate system for bistatic clutter measurements: θ_i = incident angle (in xz plane); θ_s = scattering angle (in plane containing z axis); and ϕ = out-of-plane angle (in xy plane)

Two measurement sets are of interest: in plane, where $\phi = 180°$, and out of plane, where $\phi < 180°$. When $\phi = 180°$, $\beta = |\theta_s - \theta_i|$. In the monostatic case, $\phi = 180°$, $\beta = 0$, and $\theta_s = \theta_i$. In-plane data is shown in bold type on the table. Out-of-plane data is often used in scatter jamming (*hot clutter*) calculations.

The bistatic angle is calculated from the angles in Figure 23.9 by the use of direction cosines:

$$\beta = \cos^{-1}(\cos\theta_i \cos\theta_s - \sin\theta_i \sin\theta_s \cos\phi) \tag{23.17}$$

Trends in this bistatic scattering coefficient database are summarized from Willis,[1] Weiner,[89] and Chapter 9 in Willis and Griffiths[2] as follows:

- Most of the σ_B^0 database is at X band, with 439 out of 650 data curves for both terrain and sea clutter taken by 7 of the 9 organizations. The remaining database consists of 172 data curves at L band (terrain only), 15 at S band (terrain only), 7 at C band (sea only), and 17 at K_a band (terrain only), each provided by one organization. No data is available at VHF or UHF. Thus only X band allows choices in selecting data. The Cost/Peake[90,91] and Domville[94–96] in-plane data show good correlation.[1,108]
- Attempts to model σ_B^0 data have been made using geometrical, statistical, and semi-empirical techniques, including variations of those used to model monostatic data. Meaningful results have only been achieved over a narrow range of in-plane data ($\phi = 180°$).
- Values of σ_B^0 for $\phi > \sim 140°$ are not appreciably different (within ~ 5 dB) from the monostatic case.
- Values of σ_B^0 in a broad angular region centered on $\phi = 90°$ are significantly lower than elsewhere and typically 10 to 20 dB below the monostatic value; consequently, bistatic radar surveillance can be enhanced and hot clutter can be reduced in these regions.
- Values of σ_B^0 are significantly larger near the forward-scattered, specular direction ($\phi = 0°$, $\theta_i = \theta_s$) than elsewhere and may, in some cases, reduce advantages of the enhanced, forward-scattered target RCS, particularly at frequencies > 300 MHz.
- The bistatic-monostatic equivalence theorem used to model the RCS of some targets is not generally useful for clutter modeling, except to indicate an upper limit to σ_B^0 in some regions.

TABLE 23.3 Summary of Measurement Programs for Bistatic Scattering Coefficient, σ_B^0 (In-plane data is shown in bold type—see subsequent text.) (*after M. M. Weiner, Chapter 9,*[2] *courtesy SciTech*)

Year reported[Ref]	Organization	Author	Surface Composition	# Data Curves / Figures	Frequency (GHz)	Polarization	Measurement Angles (degrees)		
							θ_i	θ_s	ϕ
1965[90]	The Ohio State University Antenna Lab.	Cost, Peake	Smooth sand		10	VV HH HV	60 – 85	60 – 85	0 – 145
			Loam				20 – 80	0 – 85	0, **180**
			Foliage, soybean						
1968[91]			Rough sand	179 / 32	10	VV HH HV	20 – 85	0 – 85	0 – **180**
			Loam with stubble						
			Grass						
1966[92]	Johns Hopkins Univ. (APL)	Pidgeon	Sea (sea states 1, 2, 3)	7 / 1	C band	VV VH	87 – 89.9	0 – 80	**180**
1967[93]			Sea (Beaufort wind 5)	1 / 1	X band	HH	82 – 89	45 – 78	**180**
1967[94]	GEC Electronics Ltd., England	Domville	Rural land		X band	VV	0 – 90	0 – 90	0, 165, **180**
			Urban land						
			Forest						
1968[95]			Sea	77 / 4	X band	VV HH	0 – 90	0 – 90	0, 165, **180**
1969[96]			Semi-desert, wet		X band	VV HH	0 – 90	0 – 90	0, 165, **180**

TABLE 23.3 Summary of Measurement Programs for Bistatic Scattering Coefficient, σ_B^0 (In-plane data is shown in bold type—see subsequent text.) (*after M. M. Weiner, Chapter 9,*[2] *courtesy SciTech*) (*Continued*)

Year reported[Ref]	Organization	Author	Surface Composition	# Data Curves / Figures	Frequency (GHz)	Polarization	Measurement Angles (degrees) θ_i	θ_s	ϕ
1977[97]	The University of Michigan (ERIM)	Larson, Heimiller, et al.	Grass with cement taxiway	16 / 8	1.3, 9.4	HH HV	50 – 80 70,75, 80 70, 75, 80	85 80 70	0 – 105 0 – **180** 0 – 105
1978[98]			Weeds and scrub trees	10 / 5	1.3, 9.4	HH HV	50 – 80	85	0 – 105
1979[99]			Orchard, weeds, scrub trees w/ snow cover	146 / 146	1.3, 9.4	HH HV	60, 70, 80 60 – 80	60 – 84 60 – 80	0 – **180** 0 – **180**
1979[100]	Raytheon Co. Wayland, MA	Cornwell, Lancaster	Beach and sand dunes Sea (sea state 2)	None	9.1	VV	Low graz. angle	Low graz. angle	0 ($\beta \cong 180°$)
1982[101] 1984[102]	Georgia Inst. of Tech. EES	Ewell, Zehner	Sea (0.9-m, 1.2–1.8-m wave heights)	7 / 7	9.38	VV HH	Low graz. angle	Low graz. angle	95 – 157
1988[103]	The University of Michigan EE / CS Dept.	Ulaby et al.	Visually smooth sand Rough sand Gravel	17 / 10	35	VV HH VH HV VH HV	66 60 60	66 60 10 – 80	0 – 170 0 – 170 0 – 90
1992[104]	MIT Lincoln Lab., MA	Kochanski	Sea (sea state 1)	3 / 3	10	VH VV	89.7	50 – 85	**180**
1994[105]	Northeastern Univ., MA	McLaughlin et al.	Forested hills	15 / 15	S band	VV HH	Low graz. angle	Low graz. angle	20 – 70
1995[106]	Northeastern Univ., MA				S band	VH HV			20 – 70
2002[107]	Univ. of MA				2.71	Fully polar			28 – 66

In addition to this database, bistatic reflectivity measurements have been made at optical[109] and sonic[110] wavelengths and of buildings,[111] airport structures,[112] and planetary surfaces.[2,52] In each of these measurements, the reflectivity data is expressed in terms of reflected power, not σ_B^0.

23.9 UNIQUE PROBLEMS AND REQUIREMENTS

In the previous edition of this book, this section covered such hardware problems as time and phase synchronization between transmitter and receiver constrained by technology available in the 1980s. Phase stability was also an issue. Since then massive advances in digital signal correlation and processing, coupled with great reductions in the cost of hardware to execute such processing, have mitigated these problems. Many recent bistatic radar programs have demonstrated quite adequate synchronization and stability—as well as detection performance—using off-the-shelf, commercial hardware. Notable examples are the NATO air defense trials[2,113,114] and the University of Washington's *Manastash Ridge Radar* measuring ionosphere turbulence[2,48]—both passive bistatic radars exploiting FM broadcast transmitters; the *HDTV-Based Passive Radar* exploiting a high definition-TV broadcast transmitter for air surveillance[2,51]; the inexpensive, commercial bistatic receiver hitchhiking off weather radars to measure full vector wind fields [2,42]; and the bistatic radar for weapons location.[35]

Furthermore, major progress has been made in developing signal and data processing algorithms, including bistatic SAR autofocus and image formation and space-time adaptive processing for bistatic airborne MTI (see Section 23.5). However, two problems continue to plague bistatic and multistatic radars and have become the topics of this section: (1) beam scan-on-scan for bistatic radars and radar hitchhikers and (2) noncooperative RF environments for passive bistatic radars. These problems and their potential remedies are detailed next.

Beam Scan-on-Scan. When high-gain, narrow-beam scanning antennas are used by both transmitter and receiver in a bistatic surveillance radar, inefficient use is made of the radar energy because only the volume common to both beams can be observed by the receiver at any given time. Outside this common beam volume, targets are lost to the receiver. Figure 23.10 shows the geometry. This problem commonly arises when attempting to hitchhike off monostatic surveillance radars. Four remedies are available to mitigate the beam scan-on-scan problem: (1) step scanning, (2) floodlight beams, (3) multiple beams, and (4) time-multiplexed beams, which in the limit is called *pulse chasing*.

Step Scanning. For hitchhiking, the step scanning remedy consists of fixing the receive beam and waiting for the transmit beam to scan through the surveillance sector. The receive beam is then stepped one beamwidth for the next transmit beam scan and so on, until the receive beam has stepped across the full surveillance sector. For a dedicated transmitter, the process can be reversed: fix the transmit beam and scan the receive beam. This remedy increases the surveillance frame time by the number of required beam steps and is usually not acceptable for large area surveillance. It can be considered in an over-the-shoulder geometry or when the baseline is small. In these cases, transmit and receive beams become more closely aligned in a

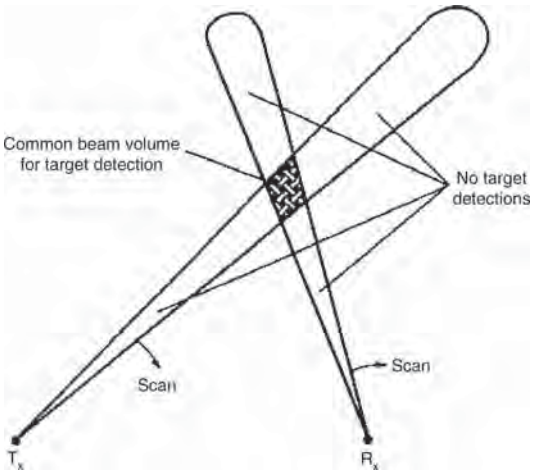


FIGURE 23.10 Beam scan-on-scan coverage problem shown in two dimensions in the bistatic plane[2] (*Courtesy SciTech*)

pseudo-monostatic geometry, which reduces the number of required beam steps. The *Bistatic Radar for Weapons Location* test program[35] is an example.

Floodlight Beams. A floodlight beam can be used with either transmitter or receiver. The floodlight transmitter remedy requires a dedicated transmit antenna that is designed to flood the surveillance sector continuously. The receiver then scans the sector with a high-gain antenna. This remedy restores the surveillance frame time lost by step scanning while simultaneously servicing multiple receivers. However, it incurs a detection range penalty by the reduced transmit antenna gain and also suffers increased sidelobe clutter levels. The floodlight receiver remedy can be used by a hitchhiker to flood a sector scanned by the transmit beam, again restoring the surveillance frame time. In addition to the range penalty, the receiver will also suffer increased clutter levels and angle measurement errors. Despite these limitations, the *Binet* floodlight receiver was found quite adequate for measuring three dimensional vector wind fields.[2,42]

Multiple Beams. A bistatic receiver can use multiple simultaneous fixed receive beams to cover the surveillance sector, which again restores the surveillance frame time. If the gain of each receive antenna is made equal to the gain of the initial single receive antenna, range performance is also restored. However, this remedy increases the cost and complexity of the receiver since a special beamforming network is required, along with a receiver and signal processor (RSP) for each beam. The multibeam receiver can be used with any type of transmitter, including a floodlight transmitter where the loss in range performance might be offset by an increase in dwell time on target, as detailed subsequently.

Time-multiplexing. If the transmitter's beam scanning schedule is known, the number of receive beams (and RSPs) can be reduced in some geometries by time-multiplexing them to cover only the currently illuminated surveillance sector.

For example, in an over-the-shoulder geometry, the receiver might use a set of beams to cover the north side of the baseline; then, as the transmit beam scans past the receiver, it switches the set to the south side, thus halving the total number required. The *Bistatic Alerting and Cueing* test program used time-multiplexed beams when hitchhiking off the AWACS transmitter for short-range air surveillance.[46,47]

Pulse Chasing. If the transmitter's beam scanning *and* pulse transmission schedule are known, pulse chasing can be considered to reduce the multibeam cost penalty further.[1,21,115–118] This was successfully demonstrated in the *Bistatic Radar for Weapons Location* test program.[35] The simplest pulse-chasing concept uses a single beam and RSP that rapidly scans the volume covered by the transmit beam, chasing the pulse as it propagates from the transmitter. The receive beam-scanning rate must be at the transmitter's pulse propagation rate, modified by the usual geometric conditions. This rate, $\dot{\theta}_R$, was initially identified by Jackson[21] and subsequently verified by Moyer and Morgan[119]:

$$\dot{\theta}_R = c \tan(\beta/2)/R_R \tag{23.18}$$

For operation in the co-site region (see Table 23.2), $\dot{\theta}_R$ can vary from 1°/μs near the baseline to 0.01°/μs when $R_T + R_R > L$. Typical $\dot{\theta}_R$ contours are shown in Jackson.[21] These rates and rate changes require an inertialess antenna, for example, a phased array with diode phase shifters. Normally, a phased array antenna used for surveillance is programmed to switch beams in increments of a beamwidth. Fractional shifts of a beamwidth can be achieved by changing the phase of a few (symmetric) pairs of phase shifters in the array. In this way, a pseudo-continuous beam scan can be generated with the required rates and rate changes.[120]

Because of pulse propagation delays from the target to the receiver, the pointing angle of the receive beam θ_R must lag the actual pulse position. For an instantaneous pulse position that generates a bistatic angle $\beta/2$, $\theta_R = \theta_T - \beta$. In terms of the bistatic triangle, the required receive beam-pointing angle is[21]

$$\theta_R = \theta_T - 2 \tan^{-1}\left(\frac{L \cos \theta_T}{R_T + R_R - L \sin \theta_T}\right) \tag{23.19}$$

The minimum receive beamwidth $(\Delta\theta_R)_m$ required to capture all returns from a range cell intersecting the common beam area is approximated by[21]

$$(\Delta\theta_R)_m \approx (c\tau_u \tan(\beta/2) + \Delta\theta_T R_T)/R_R \tag{23.20}$$

where τ_u is the uncompressed pulse width and $\Delta\theta_T$ is the transmit beamwidth. The approximation assumes that respective rays from the transmit and receive beams are parallel. The approximation is reasonable when $(R_T + R_R) \gg L$ or when $L \gg c\tau_u$. Equation 23.20 shows that $(\Delta\theta_R)_m$ changes as the receive beam scans out the transmit beam. Phased array antennas operating with a digital beamformer[59,60] can accommodate this change. Otherwise, use of a fixed beamwidth incurs a small beam mismatch loss. An example is given in Willis.[1]

Even though one pulse must be chased at a time, a hitchhiker operating in the co-site region has time to capture all pulses from a monostatic radar that uses range-unambiguous PRFs. Furthermore, when operating in the transmit- or receive-centered ovals (refer to Table 23.2) a hitchhiker can operate with range-ambiguous PRFs, for example, when transmitted from an airborne radar. Examples are given in Willis.[1]

Other implementations of pulse chasing are possible. In one concept, the fixed multibeam receive antenna is used and two RSPs are time-multiplexed across the multibeams. One RSP steps across the even-numbered beams, and the other RSP steps across the odd-numbered beams, so that returns in beam pairs are processed simultaneously: (1,2), (2,3), (3,4), etc. This *leapfrog* sequence is required to capture all returns in the common-beam area.

A second concept uses two beams and two RSPs step-scanning over the volume covered by the multibeam antenna. It uses an identical leapfrog sequence. Both concepts relax the fractional beam scan requirements by either sampling or stepping the beams in units of a beamwidth. Since they both process returns across two beamwidths before switching, the beam dwell time T_b is approximately $2(\Delta\theta_R)_m R_R / c$ and the stepping rate is T_b^{-1}. The approximation assumes negligible phase-shift delays and settling times.

Moving Target Indication (MTI) can be used with any of these pulse chasing implementations, as long as the receive beam precisely retraces its scan pattern on successive sweeps to capture the same clutter samples over the MTI processing time.

Combinations. Combinations of these remedies can be considered. For example, a fixed multibeam receive antenna can be used with a fixed floodlight transmit antenna. This configuration allows the receiver to integrate longer, subject to target/cell migration limits, which, in turn, can recover some of the lost range performance of the floodlight antenna. It also has the benefits of increasing data rates and simultaneously servicing multiple receivers. It incurs increased sidelobe clutter levels, as well as complexity and cost. Some passive bistatic radars operate in this configuration, where the floodlight transmitter is provided by a TV or FM broadcast station.[49,50]

A single receive beam can be used with a transmit antenna that is adaptively tapered to flood only the angular region covered by the receive beam at a given look angle, with the tapering such that the signal-to-noise ratio at the receiver is held constant at all positions along the receive beam. This scheme is analogous to the monostatic air surveillance radar using a cosecant-squared antenna pattern, where the echo is independent of range for a constant altitude target.[15] It has the potential of restoring much of the frame time and range performance, but incurs increased sidelobe clutter levels and increased transmitter cost and complexity. An example is given in Willis.[1]

Noncooperative RF Environment. Most passive bistatic radar (PBR) concepts and developments exploit commercial broadcast transmitters as their source of radar illumination. FM and High Definition (HD) TV terrestrial broadcast transmitters are particularly attractive due to their high-power, noise-like waveforms, and relatively wide bandwidths.[49–51] When these broadcast transmitters are appropriately sited and operating, they can support many types of surveillance, particularly air surveillance, which is often restricted for monostatic radars operating at VHF/UHF. The surveillance can be covert because even the transmitter is unaware it is being exploited and can be counter-stealth due to unavoidable aircraft resonances in the VHF/UHF region. Other attractive features are lower prime power requirements and lower costs for the PBR.[2]

While a PBR can exploit both cooperative and noncooperative broadcast transmitters, the PBR has no control over their transmission or waveform properties, specifically the transmission schedule, effective radiated power, spatial coverage, modulation type, modulation content, and resulting autocorrelation function, as outlined earlier. Furthermore, interference from the host emitter and other emitters, especially in urban and suburban areas, can significantly degrade PBR performance. This section summarizes the problems and remedies encountered by a PBR exploiting these broadcast transmitters.

Waveforms. The effective radiated power (ERP) of broadcast transmitters can vary from a maximum of ~1 MW for TV transmitters to a minimum of ~10 W for cell-phone tower transmitters. The former can yield equivalent monostatic detection ranges of air targets of 100–150 km; the latter, 1–5 km, which is of the order of the cell-phone waveform resolution, typically 2 km.[2] Consequently when these low-powered transmitters are evaluated for short-range ground or air target location.[121] only doppler (and coarse DOA) data are available, which severely restricts location capability, as outlined in Section 23.6. Thus, the low ERP of these transmitters constrained by the available bandwidth conspire to significantly reduce their utility for PBR surveillance.

The type of modulation used by a broadcast transmitter is particularly important. For example, the 1985 Crystal Palace TV transmitter trials in London[122] attempted range measurements with analog TV waveforms but found that they generated high range sidelobes (~5 dB), range ambiguities every 9.6 km, and modest range resolution (~4 km), and concluded that such waveforms were more suitable for doppler measurements. This finding established the precedence for subsequent PBR developments: doppler exploitation of stable, narrow-band carrier lines in TV transmissions and range/doppler exploitation of the wider band, noise-like spectrum of FM transmissions.

The modulation content of many broadcast transmitters changes as a function of time, thereby complicating matched filtering by the PBR's receiver. Specifically, the receiver must sample and store a segment of the direct path waveform and then cross-correlate it with the returned echo, all in real time. Since cross-correlation must be performed over the range of expected target echo time delays and/or doppler shifts, correlation receiver complexity is increased with respect to a matched-filter receiver typically used by monostatic radar.[123] Such cross-correlation is now feasible to implement, but complicates an otherwise well-developed operation in the stable, more predictable monostatic world.

A related modulation content problem is *dead air*, where no information is broadcast; thus, the broadcast transmitter modulation goes to zero and range measurement errors increase without limit. This condition can occur when broadcasting a talk show or classical music, but occurs less often with popular or rock music.[124] The frequency of this type of outage is not insignificant, occurring roughly once per second for talk broadcasts.[125] Consequently, once a track has been established a nonlinear tracking filter may be needed to edit out large error spikes (see Chapter 6 of Willis and Griffiths.[2])

Radio Frequency Interference. Passive bistatic radar performance is subject to degradation by radio frequency interference (RFI) from both the exploited broadcast transmitter and other emitters in spatial or frequency proximity. These emitters can include broadcast, communications, and navigation transmitters, as well as power tools, fluorescent lights, cooling fans, and (old) automobile ignitions, which typically generate impulsive noise.* RFI can arrive via a direct path or multipath and includes scattering from terrain or sea surfaces, also called clutter. However, the signal from an exploited transmitter arriving over the direct transmitter-to-receiver path, called the *direct-path*, is nearly always the dominant RFI source. Multipath signals from that transmitter are less severe but can also contribute to RFI.

RFI over the direct-path from the exploited transmitter, also called *direct-path breakthrough*, is common to all but the simplest, low-power CW radar. It becomes

* Sky noise consisting of sun, galactic, and atmospheric noise is another source of RFI, which can increase the receiver's noise temperature by a factor of 2–4 at frequencies below ~400 MHz.[30] However, this increase is usually orders of magnitude smaller than RFI from other sources and can be ignored.

particularly severe when the receiver is located in direct LOS of the transmitter, which must occur when surveillance of low-altitude air targets is required (Eqs. 23.5–23.8). If this direct path signal is not attenuated, the received signal becomes masked in range and often masked in doppler by sidelobes of the correlated direct path signal.

Direct-path breakthrough effects are similar to those of a spot noise jammer and can be characterized by an increase in the system noise temperature, $T_s = F_n T_o$, where F_n is the receiver noise figure and $T_o = 290$ K. Specifically, the amount of increase in T_s, and hence the amount of attenuation C_{dp} required to reduce the direct path signal to the level of T_s, is

$$C_{\text{dp}} = P_T G_T (G_R)_T \lambda^2 / (4\pi)^2 B L^2 (kT_s) \tag{23.21}$$

where P_T, G_T, λ, and k are defined with Eq. 23.2; $(G_R)_T$ is the receiving antenna power gain in the direction of the transmitter; B is the input RF bandwidth; and L is the baseline range.

For example, if the PBR exploits a typical FM broadcast transmitter located at a 50 km LOS from the receiver, $P_T G_T = 250$ kW, $\lambda = 3$ m, $B = 50$ kHz, and $L = 50$ km. Assuming a fixed receiving antenna beam tailored in elevation and covering a wide azimuth sector, which includes the transmitter site, $(G_R)_T$ might be 8 dBi. Also, assuming the ambient RFI environment described below, the noise spectral density $kT_s = -179$ dBW/Hz; thus, $C_{\text{dp}} = 88$ dB.

Combinations of earth masking, antenna shielding, spatial cancellation, and spectral cancellation can be used to achieve the required direct-path attenuation. A brute-force remedy is to physically block the transmit signal from the receiver with a shroud or structure, or if coverage allows, over-the-horizon separation. While many orders of magnitude attenuation are possible with these masking and shielding techniques, additional reduction is nearly always necessary. Howland reported a two-stage, spatial noise canceller with an adaptive M-stage lattice predictor ($M = 50$) as the first stage and an adaptive tapped delay line as the second stage, which achieved ~75 dB cancellation of the narrow-band stationary direct path signal.[113,114] This cancellation combined with masking achieved >90 dB attenuation, which satisfies the C_{dp} requirement in the above example. Howland also observed that receiver dynamic range ultimately limits the available cancellation, which, in turn, is set by the receiver's analog-to-digital converter.

Nearby emitters can significantly raise the system noise level, simply through spectral spillover from adjacent or nearby bands. In the U.S. the FCC mandates a gaussian spectral roll-off for many broadcast transmitters, which is sufficient to prevent interference in one-way home receivers from adjacent transmissions. However, it is not sufficient for two-way radar receivers, which necessarily must work much farther into the receiver noise.[126]

The severity of this problem was quantified by in situ measurements of several VHF and UHF bands in a dense urban environment.[127] Ambient VHF noise levels were found to be typically 45 dB greater than thermal noise and direct path illuminator signals some 45 dB greater still. Even with robust cancellation techniques, the residue of this unsuppressed RFI will increase the PBR's system noise figure by many tens of decibels: a 25 dB noise figure at VHF in urban and semi-rural areas is not uncommon.[126] This value translates into a noise spectral density of –179 dBW/Hz. Similar measurements have been made at UHF, with 20–25 dB noise figures being obtained using spectral cancellation by a least-squares channel estimator.[51] These noise figures are significantly higher than those obtained in the U.S. radar-designated VHF/UHF channels and are a penalty for exploiting broadcast transmitters of opportunity.

REFERENCES

1. N. J. Willis, *Bistatic Radar,* 2nd Ed., Silver Spring, MD: Technology Service Corp., 1995. Corrected and republished by Raleigh, NC: SciTech Publishing, Inc., 2005.
2. N. J. Willis and H. D. Griffiths, (eds.), *Advances in Bistatic Radar*, Raleigh NC: SciTech Publishing Inc., 2007.
3. R. C. Heimiller, J. E. Belyea, and P. G. Tomlinson, "Distributed array radar," *IEEE Trans*., vol. AES- 19, pp. 831–839, 1983.
4. B. D. Steinberg, *Principles of Aperture and Array System Design—Including Random and Adaptive Arrays*, New York: John Wiley & Sons, 1976.
5. B. D. Steinberg and E. Yadin, "Distributed airborne array concepts," *IEEE Trans*., vol. AES-18, pp. 219–226, 1982.
6. B. D. Steinberg, "High angular microwave resolution from distorted arrays," *Proc. Int. Comput. Conf*., vol. 23, 1980.
7. T. C. Cheston and J. Frank, "Phased array antennas," Chapter 7 in *Radar Handbook*, M. I. Skolnik, (ed.), 2nd Ed., New York: McGraw-Hill, 1990.
8. L. E. Merters and R. H. Tabeling, "Tracking instrumentation and accuracy on the Eastern Test Range," *IEEE Trans*., vol. SET-11, pp. 14–23, March 1965.
9. J. J. Scavullo and F. J. Paul, *Aerospace Ranges: Instrumentation*, Princeton, NJ: D. Van Nostrand Company, 1965.
10. B. D. Steinberg, et al., "First experimental results for the Valley Forge radio camera program," *Proc. IEEE*, vol. 67, pp. 1370–1371, September 1979.
11. B. D. Steinberg, "Radar imaging from a distributed array: The radio camera algorithm and experiments," *IEEE Trans*., vol. AP-29, pp. 740–748, September 1981.
12. "Handbook for NAVSPASUR System Orientation," vol. 1, Naval Space Surveillance System, Dahlgren, VA, July 1, 1976.
13. M. I. Skolnik, "An analysis of bistatic radar," *IRE Trans*., vol. ANE-8, pp. 19–27, March 1961.
14. J. M. Caspers, "Bistatic and multistatic radar," Chapter 36 in *Radar Handbook*, M. I. Skolnik, (ed.), New York: McGraw-Hill Book Company, 1970.
15. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw-Hill Book Company, 1980.
16. N. J. Willis, "Bistatic radar," Chapter 25, in *Radar Handbook*, M. I. Skolnik (ed.), 2nd Ed., New York: McGraw-Hill, 1990.
17. E. F. Ewing, "The applicability of bistatic radar to short range surveillance," in *IEE Conf. Radar 77, Publ*. 155, London, 1977, pp. 53–58.
18. E. F. Ewing and L. W. Dicken, "Some Applications of Bistatic and Multi-Bistatic Radars," in *Int. Radar Conf*., Paris, 1978, pp. 222–231.
19. A. Farina and E. Hanle, "Position accuracy in netted monostatic and bistatic radar," *IEEE Trans*., vol. AES-19, pp. 513–520, July 1983.
20. E. Hanle, "Survey of bistatic and multistatic radar," *Proc. IEE*, vol. 133, pt. F, pp. 587–595, December 1986.
21. M. C. Jackson, "The geometry of bistatic radar systems," *IEE Proc*., vol. 133, pt. F, pp. 604–612, December 1986.
22. D. E. N. Davies, "Use of bistatic radar techniques to improve resolution in the vertical plane," *IEE Electron. Lett*., vol. 4, pp. 170–171, May 3, 1968.
23. J. R. Forrest and J. G. Schoenenberger. "Totally independent bistatic radar receiver with real-time microprocessor scan correction," *IEEE Int. Radar Conf*., 1980, pp. 380–386.
24. J. G. Schoenenberger and J. R. Forrest, "Principles of independent receivers for use with co-operative radar transmitters," *Radio Electron. Eng*., vol. 52, pp. 93–101, February 1982.
25. E. G. McCall, "Bistatic clutter in a moving receiver system," *RCA Rev*., pp. 518–540, September 1969.

26. H. A. Crowder, "Ground clutter isodops for coherent bistatic radar," *IRE Nat. Conv. Rec*., pt. 5, New York, 1959, pp. 88–94.
27. L. J. Cantafio (ed.), *Space-Based Radar Handbook*, Chapter 5, Norwood, MA: Artech House, 1989.
28. A. Farina and F. A. Studer, *Radar Data Processing*, Vol. 2, *Advanced Topics and Applications*, UK: Research Studies Press Ltd., 1986.
29. I. Stein, "Bistatic radar applications in passive systems," *Journal of Electronic Defense*, 13(3), pp. 55–61, March 1990.
30. D. K. Barton, *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988.
31. D. K. Barton, private communication, June 2006.
32. R. L. Easton and J. J. Fleming, "The Navy space surveillance system," *Proc. IRE*, vol. 48, pp. 663– 669, 1960.
33. R. J. Lefevre, "Bistatic radar: New application for an old technique," *WESCON Conf. Rec*., San Francisco, 1979, pp. 1–20.
34. F. L. Fleming and N. J. Willis, "Sanctuary radar," *Proc. Mil. Microwaves Conf*., London, October 22–24, 1980, pp. 103–108.
35. L. Bovino, "Bistatic radar for weapons location," U.S. Army Communications Electronics Command, Fort Monmouth, NJ, 1994.
36. Russia's Arms Catalog, vol. 5, Air Defense, Moscow: Military Parade Ltd., 1997.
37. "Barrier," Bistatical low flying target detection system, Nizhny Novgorod Scientific-Research Radiotechnical Institute, Moscow, 2000.
38. A. G. Blyakhman et al., "Forward scattering radar moving object coordinate measurement," *IEEE International Radar Conference*, 2000.
39. A. Thomson, *A GRAVES Sourcebook,* version of 2006-10-27, thomsona@flash.net.
40. J. E. Salah and J. E. Morriello, "Development of a multistatic measurement system," in *IEEE International Radar Conference*, 1980, pp. 88–93.
41. "Multistatic mode raises radar accuracy," *Aviation Week and Space Technology*, pp. 62–69, July 14, 1980.
42. J. Wurman, S. Heckman, and D. Boccippio, "A bistatic multiple-doppler radar network,"*Journal of Applied Meterology*, vol. 32, pp. 1802–1814, December 1993.
43. J. Wurman, M. Randall, C. L. Frush, E. Loew, and C. L. Holloway, "Design of a bistatic dual-doppler radar for retrieving vector winds using one transmitter and a remote low-gain passive receiver," invited paper, *Proc. of the IEEE*, vol. 82, no. 12, December 1994, pp. 1861–1872.
44. F. Johnson, *Synthetic Aperture Radar (SAR) Heritage: An Air Force Perspective*, Ohio: Air Force Avionics Laboratory, Wright Patterson AFB, June 2003.
45. D. C. Lorti and M. Balser, "Simulated performance of a tactical bistatic radar system," *IEEE EASCON 77 Rec. Publ*. 77 CH1255-9, Arlington, VA, 1977, pp. 4-4A–4-40.
46. "Bistatic Radars Hold Promise for Future System," *Microwave Systems News*, pp. 119–136, October 1984.
47. E. C. Thompson, "Bistatic radar noncooperative illuminator synchronization techniques," in *Proc. of the 1989 IEEE National Radar Conference*, Dallas, TX, March 29–30, 1989.
48. J. D. Sahr, "Remote sensing with passive radar at the University of Washington," *IEEE Geoscience and Remote Sensing Society Newsletter*, pp. 16–21, December 2005.
49. "Passive system hints at stealth detection silent sentry—A new type of radar," *Aviation Week and Space Technology*, November 30, 1998, pp. 70–71.
50. J. Baniak, G. Baker, A. M. Cunningham, and L. Martin, "Silent Sentry™ Passive Surveillance," *Aviation Week and Space Technology*, June 7, 1999.
51. A. Andrews, "HDTV-based passive radar," presented at AOC 4th Multinational PCR Conference, Syracuse, NY, October 6, 2005.
52. R. A. Simpson, "Spacecraft studies of planetary surfaces using bistatic radar," *IEEE Trans. Geoscience and Remote Sensing*, vol. 31 no. 2, March 1993.

53. D. Prichard, *The Radar War: The German Achievement, 1904–1945*, Cambridge, UK: Patrick Stephens Ltd., 1989.
54. A. Price, *Instruments of Darkness: The History of Electronic Warfare*, New York: Charles Scribner's Sons, 1978.
55. P. J. Klass, "Navy improves accuracy, detection range," *Aviation Week and Space Technology*, pp. 56–61, August 16, 1965.
56. *Technology in the National Interest*, Lexington, MA: MIT Lincoln Laboratory, 1995.
57. A. Bernard, private communication, MIT Lincoln Laboratory, July 24, 2006.
58. S. Satoh and J. Wurman, "Accuracy of wind fields observed by a bistatic doppler radar network," *Journal Ocean. Atmos. Tech.* vol. 20, pp. 1077–1091, 2003.
59. A. E. Ruvin and L. Weinberg, "Digital multiple beamforming techniques for radars," in *IEEE Eascon '78 Rec.*, pp. 152–163.
60. E. E. Swartzlander and J. M. McKay, "A digital beamforming processor," *Real Time Signal Processing III, SPIE Proc*., vol. 241, pp. 232–237, 1980.
61. A. Farina, "Tracking function in bistatic and multistatic radar systems," *Proc. IEE*, vol. 133, pt. F, pp. 630–637, December 1986.
62. R. B. Patton, Jr., "Orbit determination from single pass doppler observations," *IRE Transactions on Military Electronics*, pp. 336–344, April–July, 1960.
63. A. Farina, "Tracking function in bistatic and multistatic radar systems," *IEE Proc*., 133(7), pt. F, pp. 630–637, December 1986.
64. M. I. Skolnik, *Introduction to Radar Systems*, New York: McGraw Hill Book Co., 1962.
65. D. K. Barton, *Radar System Analysis*, Dedham, MA: Artech House, Inc., 1976.
66. K. M. Siegel, et al., "Bistatic radar cross sections of surfaces of revolution," *J. Appl. Phys*., vol. 26, pp. 297–305, March 1955.
67. K. M. Siegel, "Bistatic radars and forward scattering," in *Proc. Nat. Conf. Aeronaut. Electron*., May 12–14, 1958, pp. 286–290.
68. F. V. Schultz et al., "Measurement of the radar cross-section of a man," *Proc. IRE*, vol. 46, pp. 476–481, February 1958.
69. J. W. Crispin, Jr. et al., "A theoretical method for the calculation of radar cross section of aircraft and missies," University of Michigan, Radiation Lab. Rept. 2591-1-H, July 1959.
70. R. E. Hiatt et al., "Forward scattering by coated objects illuminated by short wavelength radar," *Proc. IRE*, vol. 48, pp. 1630–1635, September 1960.
71. R. J. Garbacz and D. L. Moffett, "An experimental study of bistatic scattering from some small, absorber-coated, metal shapes," *Proc. IRE*, vol.49, pp. 1184–1192, July 1961.
72. M. G. Andreasen, "Scattering from bodies of revolution," *IEEE Trans*., vol. AP-13, pp. 303–310, March 1965.
73. C. R. Mullin et al., "A numerical technique for the determination of the scattering cross sections of infinite cylinders of arbitrary geometric cross section," *IEEE Trans*., vol. AP-13, pp. 141–149, January 1965.
74. R. E. Kell, "On the derivation of bistatic RCS from monostatic measurements," *Proc. IEEE*, vol. 53, pp. 983–988, August 1965.
75. W. I. Kock, "Related experiments with sound waves and electromagnetic waves," *Proc. IRE*, vol. 47, pp. 1200–1201, July 1959.
76. K. M. Siegel et al., "RCS calculation of simple shapes—bistatic," Chapter 5 in *Methods of Radar Cross-Section Analysis*, New York: Academic Press, 1968.
77. H. Weil et al., "Scattering of electromagnetic waves by spheres," University of Michigan, Radiat. Lab. Stud. Radar Cross Sections X, Rept. 2255-20-T, contract AF 30(602)-1070, July 1956.
78. R. W. P. King and T. T. Wu, *The Scattering and Diffraction of Waves*, Cambridge, MA: Harvard Universities Press, 1959.
79. R. F. Goodrich et al., "Diffraction and scattering by regular bodies—I: The sphere," University of Michigan, Dept. Electr. Eng. Rept. 3648-1-T, 1961.

80. M. Matsuo et al., "Bistatic radar cross section measurements by pendulum method," *IEEE Trans*., vol. AP-18, pp. 83–88, January 1970.
81. G. W. Ewell and S. P. Zehner, "Bistatic radar cross section of ship targets," *IEEE J. Ocean. Eng*., vol. OE-5, pp. 211–215, October 1980.
82. "Radar cross-section measurements," General Motors Corporation, Delco Electron. Div. Rept. R81-152, Santa Barbara, CA, 1981.
83. C. G. Bachman, *Radar Targets*, Lexington, MA: Lexington Books, 1982, p. 29.
84. F. C. Paddison et al., "Large bistatic angle radar cross section of a right circular cylinder," *Electromagnetics*, vol. 5, pp. 63–77, 1985.
85. J. I. Glaser, "Bistatic RCS of complex objects near forward scatter," *IEEE Trans*., vol. AES-21, pp. 70–78, January 1985.
86. C.-C. Cha et al., "An RCS analysis of generic airborne vehicles' dependence on frequency and bistatic angle," in *IEEE Nat. Radar Conf*., Ann Arbor, MI, April 20, 1988, pp. 214–219.
87. W. A. Pierson et al., "The effect of coupling on monostatic-bistatic equivalence," *Proc. IEEE*, pp. 84–86, January 1971.
88. "Bistatic radars hold promise for future systems," *Microwave Syst. News*, pp. 119–136, October 1984.
89. M. M. Weiner, "Multistatic radar phenomenology terrain & sea scatter," RADC-TR-81-75, vol. 1, May 1981, now unlimited distribution.
90. S. T. Cost, "Measurements of the bistatic echo area of terrain at X-band," The Ohio State University, Antenna Laboratory, Report No. 1822-2, May 1965.
91. W. H. Peake and S.T. Cost, "The bistatic echo area of terrain at 10 GHz," in *IEEE WESCON* 1968, Session 22/2, pp. 1–10.
92. V. W. Pidgeon, "Bistatic cross section of the sea," *IEEE Trans*. AP-14(3), pp. 405–406, May 1966.
93. V. W. Pidgeon, "Bistatic cross section of the sea for Beauford 5 sea," in *Science and Technology*, vol. 17 "Use of space systems for planetary geology and geophysics," San Diego: American Astronautical Society, 1968, pp. 447–448.
94. A. R. Domville, "The bistatic reflection from land and sea of X-band radio waves, Part I," GEC (Electronics) Ltd., Stanmore, England, Memorandum SLM1802, July 1967.
95. A. R. Domville, "The bistatic reflection from land and sea of X-band radio waves, Part II," GEC (Electronics) Ltd., Stanmore, England, Memorandum SLM2116, July 1968.
96. A. R. Domville, "The bistatic reflection from land and sea of X-band radio waves, Part II-supplement," GEC (Electronics) Ltd., Stanmore, England, Memorandum SLM2116 (Supplement), July 1969.
97. R. W. Larsen and R. C. Heimiller, "Bistatic clutter data measurement program," Environmental Research Institute of Michigan, RADC-TR-77-389, November 1977, AD-A049037.
98. R. W. Larsen, A. L. Maffett, R. C. Heimiller, A. F. Fromm, E. L. Johansen, R. F. Rawson, and F. L. Smith, "Bistatic clutter measurements," IEEE Trans. AP-26(6), pp. 801–804, November 1978.
99. R. W. Larsen, A. Maffett, F. Smith, R.C. Heimiller, and A. Fromm, "Measurements of bistatic clutter cross section," Environmental Research Institute of Michigan, Final Technical Report RADC-TR-79-15, May 1979, AD-A071193.
100. P. E. Cornwell and J. Lancaster, "Low-altitude tracking over rough surfaces II: Experimental and model comparisons," in *IEEE EASCON-79 Record*, October 1979, pp. 235–248.
101. G. W. Ewell and S. P. Zehner, "Bistatic sea clutter return near grazing incidence," in *IEE Int. Conf. Radar 82*, Publication No. 216, London, October 1982, pp. 188–192.
102. G. W. Ewell, "Techniques of radar reflectivity measurement," Chapter 7 in *Bistatic Radar Cross-Section Measurements*, 2nd Ed., N. C. Currie (ed.), Norwood, MA: Artech House,1984.
103. F. T. Ulaby, T. E. Van Deventer, J. R. East, T. F. Haddock, and M. E. Coluzzi, "Millimeter-wave bistatic scattering from ground and vegetation targets," *IEEE Trans*. GRS-26(3), pp. 229–243, May 1988.

104. T. P. Kochanski, M. J. Vanderhill, J.V. Zolotarevsky, and T. Fariss, "Low illumination angle bistatic sea clutter measurements at X-band," in *IEEE Int. Conf. Oceans-92: Mastering the Oceans Through Technology, Proc*., vol. 1, October 26–29, 1992, pp. 518–523.
105. D. J. McLaughlin, E. Boltniew, Y. Wu, and R. S. Raghavan, "Low grazing angle bistatic NCRS of forested clutter," *Electronics Letters*-30(18), pp. 1532–1533, September 1, 1994.
106. D. J. McLaughlin, E. Boltniew, R. S. Raghavan, and M. J. Sowa, "Cross-polarized bistatic clutter measurements," *Electronics Letters*-31(6), pp. 490–491, March 16, 1995.
107. D. J. McLaughlin, Y. Wu, W. G. Stevens, X. Zhang, M. J. Sowa, and B. Weijers, "Fully polarimetric bistatic radar scattering behavior of forested hills," *IEEE Trans*. AP-5O(2), pp. 101–110, February 2002.
108. R. E. Vander Schurr and P. G. Tomlinson, "Bistatic clutter analysis," Decision-Sciences Applications, Inc., RADC-TR-79-70, April 1979.
109. G. O. Sauermann and P. C. Waterman, "Scattering modeling: Investigation of scattering by rough surfaces," MITRE Corporation, Rept. MTR-2762, *AFAL-TR*-73-334, January 1974.
110. J. G. Zornig et al., "Bistatic surface scattering strength at short wavelengths," Yale University, Dept. Eng. Appl. Sci. Rept. *CS-9*, *AD-A*041316, June 1977.
111. E. N. Bramley and S. M. Cherry, "Investigation of microwave scattering by tall buildings," *Proc. IEE*, vol. 120, pp. 833–842, August 1973.
112. A. E. Brindly et al., "A Joint Army/Air Force investigation of reflection coefficient at *C* and K_u bands for vertical, horizontal and circular system polarizations," IIT Research Institute, Final Rept., *TR*-76-67, *AD-A*031403, Chicago, IL, July 1976.
113. P. E. Howland, D. Maksimiuk, and G. Reitsma, "FM radio based bistatic radar," *IEE Proc.-Radar Sonar Navig*., vol. 152, no. 3, pp. 107–115, June 2005.
114. P. E. Howland, "FM-radio based bistatic radar," presented at AOC 4th Multinational PCR Conference, Syracuse, NY, October 6, 2005.
115. T. A. Soame and D. M. Gould, "Description of an experimental bistatic radar system," in *IEE Int. Radar Conf. Publ.* 281, 1987, pp. 12–16.
116. E. Hanle, "Pulse chasing with bistatic radar-combined space-time filtering," in *Signal Processing II: Theories and Applications*, H. W. Schussler (ed.), North Holland: Elsevier Science Publishers B. V., pp. 665–668.
117. J. G. Schoenenberger and J. R. Forrest, "Principles of independent receivers for use with co-operative radar transmitters," *Radio Electron. Eng*., vol. 52, pp. 93–101, February 1982.
118. N. Freedman, "Bistatic radar system configuration and evaluation," Raytheon Company, Independ. Dev. Proj. 76*D*-220, *Final Rept*. *ER*76-4414, December 30, 1976.
119. L. R. Moyer, "Comments on 'Receiver antenna scan rate requirements needed to implement pulse chasing in a bistatic radar receiver,'" *IEEE Trans. on Aerospace and Electronic Systems*, vol. 38, no. 1, p. 300, January 2002, correspondence.
120. J. Frank and J. Ruze, "Beam steering increments for a phased array," *IEEE Trans.*, vol. AP-15, pp. 820–821, November 1967.
121. D. K. P. Tan, H. Sun, Y. Lu, M. Lesturgie, and H. L. Chan, "Passive radar using global system for mobile communication signal: Theory, implementation and measurements," *IEE Proc.-Radar Sonar Navig*., vol. 152, no. 3, June 2005.
122. H. D. Griffiths and N. R. W. Long, "Television-based bistatic radar," *IEE Proc*., 133(7), Pt. F, pp. 649–657, December 1986.
123. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2001.
124. H. D. Griffiths et al., "Measurement and analysis of ambiguity functions of off-air signals for passive coherent location," *Electronics Letters*, vol. 39, no. 13, June 26, 2003.
125. M. A. Ringer and G. J. Glazer, "Waveform analysis of transmissions of opportunity for passive radar," in *Fifth International Symposium on Signal Processing and its Applications*, Brisbane, Australia, August 1999, pp. 511–514.
126. Richard Lodwig, private communication, Lockheed Martin Mission Systems, 2003.
127. H. D. Griffiths and C. J. Baker, "Passive coherent location radar systems. Part 1: Performance prediction," *IEE Proc.-Radar Sonar Navig*., vol. 152, no. 3, pp. 153–159, June 2005.

Chapter 24
Electronic Counter-Countermeasures

A. Farina

Analysis of Integrated Systems
SELEX Sistemi Integrati

24.1 INTRODUCTION

Since World War II, both radar and electronic warfare (EW)* have achieved a very high state of performance.[1,2] Modern military forces depend heavily on electromagnetic (EM) systems for surveillance, weapon control, communication, and navigation; thus access to, and control of, the EM spectrum is vital. Electronic countermeasures (ECM) are likely to be taken by hostile forces to degrade the effectiveness of EM systems.[3–7] As a direct consequence, EM systems are more and more frequently equipped with so-called electronic counter-countermeasures (ECCM) to ensure effective use of the EM spectrum despite an enemy's use of EW actions.

This chapter is devoted to the description of the ECCM techniques and design principles to be used in radar systems when they are subject to an ECM threat. Section 24.2 starts with a recall of the definitions pertaining to EW and ECCM. The topic of radar signals interception by EW devices is introduced in Section 24.3; the first strategy to be adopted by radar designers is to try to avoid interception by the opponent electronic devices. Section 24.4 is dedicated entirely to the analysis of the major ECM techniques and strategies. It is important to understand the ECM threat to a radar system in order to be able to efficiently react to it. To facilitate the description of the crowded family of ECCM techniques (Sections 24.6 through 24.10), a classification is attempted in Section 24.5. Then, the techniques are introduced according to their use in the various sections of radar, namely, antenna, transmitter, receiver, and signal processing. A key role is also played by those ECCM techniques that cannot be classified as electronic, such as human factors, methods of radar operation, and radar deployment tactics (Section 24.10).

The ensuing Section 24.11 shows the application of the aforementioned techniques to the most common radar families, namely, surveillance, tracking, multifunctional, phased-array, imaging, and over-the-horizon radars. The main design principles (e.g., selection of transmitter power, frequency, waveform, and antenna gain) as dictated by the ECM threat are also discussed in some detail.

* A list of acronyms is at the end of the chapter before the list of references.

The chapter ends with an approach to the problem of evaluating the efficacy of ECCM and ECM techniques (Section 24.12). There is a lack of theory to properly quantify the endless battle between ECCM and ECM techniques. Nevertheless, a commonly adopted approach to determine the ECM effect on a radar system is based on evaluation of the radar range under jamming conditions. The advantage of using specific ECCM techniques can be taken into account by calculating the radar range recovery.

A list of acronyms used in the chapter and the references appear at the end of the chapter.

24.2 TERMINOLOGY

EW is defined as a military action involving the use of EM energy to determine, exploit, reduce, or prevent radar use of the EM spectrum.[8–11] The operational employment of EW relies upon the capture of radar EM emissions using electronic intelligence (ELINT) devices, collating the information in support databases that are then used to interpret EM emission data, to understand the radar system functions, and to program reactions against the radar. EW is organized into two major categories: electronic warfare support measures (ESM) and ECM. Basically, the EW community takes as its job the degradation of radar capability. The radar community takes as its job the successful application of radar in spite of what the EW community does; the goal is pursued by means of ECCM techniques. The definitions of ESM, ECM, and ECCM are listed below.[8,11,12†]

ESM is that division of EW involving actions taken to search for, intercept, locate, record, and analyze radiated EM energy for the purpose of exploiting such radiations in the support of military operations. Thus, ESM provides a source of EW information required to conduct ECM, threat detection, warning, and avoidance. ECM is that division of EW involving actions taken to prevent or reduce a radar's effective use of the EM spectrum. ECCM comprises those radar actions taken to ensure effective use of the EM spectrum despite the enemy's use of EW.

The topic of EW is extremely rich in terms, some of which are also in general use in other electronic fields. A complete glossary of terms in use in the ECM and ECCM fields is found in the literature.[8,11,13]

24.3 ELECTRONIC WARFARE SUPPORT MEASURES

ESM usually consists of several detection and measurement receivers and real time processor boards dedicated to the interception of radar emissions. The identification of specific emitters is based on comparison with tactical or strategic ELINT.[9,14–17] Emitter location can be additionally provided through several methods such as triangulation

† Since the publication of the second edition of this Handbook, the U.S. Air Force changed some of the EW terms we have gotten used to over these many years. ECM is now Electronic Attack (EA), ECCM is Electronic Protection (EP), and ESM is Electronic Support (ES). These terms are not used in this chapter because they are seldom used by the radar community who seem to prefer retaining the more familiar expressions ECM, ECCM, and ESM.

from remote systems or single platform sequential bearing measurements, difference time of arrival (DToA) or hyperbolation, and Phase Difference Rate (PDR). Modern digital receiver technology, coupled with state-of-the-art deinterleaving signal processing to cleanly isolate and identify individual EM emitters, will enhance situation awareness. Using techniques such as time and frequency difference of arrival will improve single and multiple platform spatial location; this will allow EW to be used to cue targeting systems.

Radar interception, which is of particular interest in this section, is based on the reception and measurements of the signals transmitted by radar systems whether pulse or Continuous Wave (CW). The operational scenario in which ESM should operate is generally crowded with pulsed radar signals: figures of 500,000 to 1 million pulses per second are frequently quoted in the literature.[9] The center frequency, amplitude, pulse width, time of arrival (ToA), and bearing of each detected pulse are measured, converted in digital format, and packed into a pulse descriptor word (PDW). The train of PDWs are sent to a pulse-sort processor that deinterleaves the sequences belonging to different emitters and identifies Pulse Repetition Interval (PRI) values and modulation laws (random jitter, stagger, switching). Further comparison against an emitter database, which contains the range of characteristic parameters (frequency, pulse width, PRI), the related pattern of agility (random, stagger, etc.) for each emitter, the type of antenna scanning pattern and periods permits the generation of an emitter list with an identification score. The ESM receiver is used to control the deployment and operation of ECM; the link between ESM and ECM is often automatic.

A single received radar pulse is characterized by a number of measurable parameters. The availability, resolution, and accuracy of these measurements must all be taken into account when designing the deinterleaving system because the approach used depends on the parameter data set available. Obviously, the better the resolution and accuracy of any parameter measurement, the more efficiently the pulse-sort processor can carry out its task. However, there are limitations on the measurement process from outside the ESM system (e.g., multipath), from inside the system (e.g., timing constraints, dead time during reception), and from cost-effectiveness considerations. Angle of arrival is the most important sorting parameter available to the deinterleaving process since the target bearing does not vary from pulse to pulse. Therefore, amplitude comparison monopulse antennas or multiple base interferometric (phase comparison) systems are often used in order to warrant both 360° spatial coverage and pulse-based angle of arrival measurement. Monopulse rotating antennas can also be used when the time to intercept is not critical (this is the ELINT case), and it is possible to scan sequentially the operational scenario.

The carrier frequency is the next most important pulse parameter for deinterleaving. A common method of frequency measurement is to use a scanning superheterodyne receiver that has the advantage of high sensitivity, good frequency resolution, and immunity with respect to the interference of nearby emitters.[9] Unfortunately, this type of receiver has a poor probability of intercept for the same reasons as the rotating bearing measurement system. The situation is much worse if the emitter is also frequency-agile (random variation) or frequency-hopping (systematic variation). A common method to allow for wideband frequency measurements is based on interferometric devices that provide instantaneous frequency measurement with good accuracy and are able to reject signal interference with lower intensity. The higher sensitivity and probability of interception are provided by wide instantaneous band superheterodyne receivers followed by banks of contiguous receiver channels.

Several technologies have been proposed in the past such as surface acoustic wave (SAW) filters and Bragg cells.[9] The preferred approach is based on digital receivers that integrate wideband spectral analysis and several post-detection functions, such as intrapulse modulation measurement and waveform code reconnaissance.

Pulse width is an unreliable sorting parameter because of the high degree of corruption resulting from multipath transmission. Multipath effects can severely distort the pulse envelope, for example, by creating a long tail to the pulse and even displacing the position of the peak.

The ToA of the pulse can be taken as the instant that a threshold is crossed, but in the presence of noise and distortion, this becomes a variable measurement. Nevertheless, the ToA is used for deriving the PRI of the radar. The amplitude of the pulse is taken as the peak value. Dynamic-range considerations must take into account at least some three orders of magnitude for range variation and three orders of magnitude for scan pattern variations. In practice, 60 dB instantaneous dynamic range sounds like a minimum value; in many applications, it should be larger. The amplitude measurement is used (along with ToA) for deriving the scan pattern of the emitter.[9]

The classification of radar interception systems is based on the type of representation they provide of the electronic environment. A radar warning receiver (RWR) in an airborne installation provides alerts of the presence of threats such as radar on a missile, supplying the relative bearing on a cockpit-based display. Search radars are not the primary target for these systems, though range advantage due to one-way propagation with respect to two-way propagation allows radar interception at farther range than own platform detection. Required sensitivity values range from –38 dBm (dB milliWatt with respect to the isotropic) to –60 dBm. ESM is the most complex system and usually comprises the capability to produce a picture of the complete electronic order of battle in its deployment area and alert function. This kind of system is able to detect and analyze emitter waveforms and scanning patterns. The reaction time for the reconnaissance of the operational environment may be less than 10 s, though dangerous emitters and alert functions call for tighter constraints. Required sensitivity ranges from –55 dBm to better than –80 dBm. ELINT systems are similar to ESM, but may not require 100% probability of intercept. The reaction time may be minutes or hours. The purpose is not to detect emitters as soon as they switch on in the operational environment, but to provide detailed characteristics of emitters to allow the generation of an identification database for RWR and ESM systems. ELINT system sensitivity may reach –90 dBm, but they don't need to provide 360° surveillance, and they can reach such performance by means of several directive antennas.

The range at which a radar emission is detected by an RWR depends primarily on the sensitivity of the receiver and the radiated power of the victim radar. The calculation of the warning range can be obtained by the basic *one-way beacon equation*, which provides the signal-to-noise ratio (SNR) at the RWR:

$$\left(\frac{S}{N}\right)_{\text{at RWR}} = \left(\frac{P}{4\pi R^2}\right) G_t \left(\frac{G_r \lambda^2}{4\pi}\right) \left(\frac{1}{kT_S B}\right) \frac{1}{L} \tag{24.1}$$

where P is the radar radiated power; R is the range from the RWR to the radar; G_t is the transmitting-antenna gain of the radar; G_r is the receiving-antenna gain of the RWR; λ is the radar wavelength; the quantity kT_sB is the total system noise power of the RWR; and L is the losses.

Equation 24.1 is the basis of performance calculation for an RWR. It is noted that the RWR detection performance is inversely proportional to R^2 rather than to R^4 of the radar target detection equation. For this reason, the RWR can detect a radiating radar at distances far beyond those of a radar's own target detection capability. The radar-versus-interceptor problem is a battle in which the radar's advantage lies in the use of matched filtering, which cannot be duplicated by the interceptor (it does not know the exact radar waveform), while the interceptor's advantage lies in the fundamental R^2 advantage of one-way versus two-way radar propagation.[15–18] Low probability of intercept (LPI) techniques are applied to radar to win the battle of "to see and not to be seen": see Schleher[19] and references therein.

24.4 ELECTRONIC COUNTERMEASURES

The objectives of an ECM system are to deny information (detection, position, track initiation, track update, and classification of one or more targets) that the radar seeks or to surround desired radar echoes with so many false targets that the true information cannot be extracted.[3–7]

ECM tactics and techniques may be classified in a number of ways, i.e., by main purpose, whether active or passive, by deployment, by platform, by victim radar, or by a combination of these.[13–16,20] An encyclopedia of ECM tactics and techniques can be found in the literature.[3,13] Here description is limited to the most common types of ECM.

ECM includes both jamming and deception. *Jamming* is the intentional and deliberate transmission or retransmission of amplitude, frequency, phase, or otherwise modulated intermittent, CW, or noise-like signals for the purpose of interfering with, disturbing, exploiting, deceiving, masking, or otherwise degrading the reception of other signals that are used by radar systems.[3–13] A jammer is any ECM device that transmits a signal of any duty cycle for the sole or partial purpose of jamming a radar system.[3–13]

Radio signals by special transmitters intended for interfering with or precluding the normal operation of a victim radar system are called *active jamming*. They produce at the input of a victim system a background that impedes the detection and recognition of useful signals and determination of their parameters. The most common forms of active noise jamming are spot, swept, and barrage noises. Spot noise is used when the center frequency and bandwidth of the victim system to be jammed are known and confined to a narrow band. However, many radars are frequency-agile over a wideband as an ECCM against spot jamming. If the rate of frequency agility is slow enough, the jammer can follow the frequency changes and maintain the effect of spot jamming. Barrage or broadband jamming is simultaneously radiated across the entire band of the radar spectrum of interest. This method is used against frequency-agile systems whose rates are too fast to follow or when the victim's frequency parameters are imprecisely known.

Jammer *size* is characterized by the *effective radiated power*; ERP = G_jP_j, where G_j is the transmit antenna gain of the jammer and P_j is the jammer power.

Passive ECM is synonymous with chaff, decoys, and other reflectors that require no prime power. The chaff is made of elemental passive reflectors that can be floated or otherwise suspended in the atmosphere or exoatmosphere for the purpose of confusing, screening, or otherwise adversely affecting the victim electronic system. Examples are metal foils, metal-coated dielectrics (aluminum, silver, or zinc over fiberglass or nylon

being the most common), string balls, rope, and semiconductors.[3,13] Chaff consists of dipoles cut to approximately a half wavelength of the radar frequency. It is suitably packaged to contain a broad range of dipole lengths designed to be effective over a wide frequency band. The basic properties of chaff are effective scatter area, the character and time of development of a chaff cloud, the spectra of the signals reflected by the cloud, and the width of the band that conceals the target.[3,9,21,22] From a radar viewpoint, the properties of chaff are very similar to those of weather clutter, except that its broadband in frequency can extend down to VHF. The mean doppler frequency of the chaff spectrum is determined by the mean wind velocity, while the spectrum spread is determined by wind turbulence and a shearing effect due to different wind velocities as a function of altitude.[3]

Decoys, which are another type of passive ECM, are a class of physically small radar targets whose radar cross sections (RCS) are generally enhanced by using reflectors or a Luneburg lens to simulate fighter or bomber aircraft. The objective of decoys is to cause a dilution of the assets of the defensive system, thereby increasing the survivability of the penetrating aircraft. However, when the decoys grow too large, they have to be engaged if they are thought large enough to carry a weapon.

Penetration Aid (Penaids) could be used by incoming ballistic missiles (BMs).[23] Penaid decoys are only one of several possible penaids. A decoy provides another target that the defense has to handle if the defense cannot distinguish a decoy from a re-entry vehicle.

The other major type of active jammer is deceptive ECM (DECM). *Deception* is the intentional and deliberate transmission or retransmission of amplitude, frequency, phase, or otherwise modulated intermittent or CW signals for the purpose of misleading in the interpretation or use of information by electronic systems.[3,13] The categories of deception are manipulative and imitative. *Manipulative* implies the alteration of friendly EM signals to accomplish deception, whereas *imitative* consists of introducing radiation into radar channels that imitates a hostile emission. DECM is also divided into *transponders* and *repeaters*.[3] Transponders generate noncoherent signals that emulate the temporal characteristics of the actual radar return. Repeaters generate coherent returns that attempt to emulate the amplitude, frequency, and temporal characteristics of the actual radar return. Repeaters usually require some form of memory for microwave signals to allow anticipatory returns to be generated; this is usually implemented by using a microwave acoustic memory or a digital RF memory (DRFM).[3]

In a DRFM system, the input RF signal is generally first down-shifted in frequency and then sampled with a high-speed analogue-to-digital converter (ADC). The samples, stored in memory, can be manipulated in amplitude, frequency, and phase to generate a wide range of jamming signals. The stored samples are later recalled, processed by the digital-to-analogue converter (DAC), up converted, and transmitted back to the victim radar.[24] The information content of an intercepted radar signal is mainly carried in the phase of the signal, and then the amplitude is usually discarded and only the phase contribution is quantized and processed.[25] The phase quantization is performed by the DRFM by means of M bits into $N = 2^M$ levels. After the phase quantization, introduced on the signal by the DRFM, the jamming signal is transmitted back to the victim radar with an increasing delay with respect to the received radar signal. This delay is quantized by a range gate pull off (RGPO) device. A range gate stealer system linearly delays the quantized signal in order to generate a constant range-rate false target. The joint effect of phase and delay quantization in DRFM can be analyzed as reported in Greco, Gini, and Farina.[26] Other artifacts in the deception signals can be introduced by imperfections in the down-up conversion and demodulation/modulation of the signal performed in the DRFM device. A detailed analysis of this kind of errors is in Berger.[27]

The DRFM is the principal means to implement a deception jammer; the range-gate stealer pulls the radar-tracking gate from the target position through the introduction of a false target into the radar's range-tracking circuits. A repeater jammer sends back an amplified version of the signal received from the radar. The deception signal, being stronger than the radar's return signal, captures the range-tracking circuits. The deception signal is then progressively delayed by using the DRFM, thereby "walking" the range gate off the actual target (RGPO technique). When the range gate is sufficiently removed from the actual target, the deception jammer is turned off, forcing the tracking radar into a target reacquisition mode.[3] Another form of deception is the velocity gate pull off (VGPO); a combination of RGPO and VGPO is also possible.

Another DECM technique is called *inverse-gain jamming*; it is used to capture the angle-tracking circuits of a conical-scan tracking radar.[3,13] This technique repeats a replica of the received signal with an induced amplitude modulation that is the inverse of the victim radar's combined transmitting and receiving antenna scan patterns. Against a conically scanning tracking radar, an inverse-gain repeater jammer has the effect of causing positive feedback, which pushes the tracking-radar antenna away from the target rather than toward the target. Inverse-gain jamming and RGPO are combined in many cases to counter conical-scan tracking radars.[3] The vulnerability of conical scan to such countermeasures motivates the use of monopulse trackers that are almost always used in military tracking radars.

A different form of DECM used against the main beam of surveillance radar attempts to cover the target's skin return with a wide pulse in order to confuse the radar's signal-processing circuitry into suppressing the actual target return.

What the radar can do against DECM is discussed later; see Section 24.11.

In the deployment of ECM, several classes can be singled out.[3] In the *stand-off jammer* (SOJ) case, the jamming platform remains close to but outside the lethal range of enemy weapon systems and jams these systems to protect the attacking vehicles. Stand-off ECM systems employ high-power noise jamming that must penetrate through the radar antenna receiving sidelobes at long ranges. *Escort jamming* is another ECM tactic in which the jamming platform accompanies the strike vehicles and jams radars to protect the strike vehicles.

Mutual-support, or *cooperative*, ECM involves the coordinated conduct of ECM by combat elements against acquisition and weapon control radars. One advantage of mutual-support jamming is the greater ERP available from a collection of platforms in contrast with a single platform. However, the real value of mutual-support jamming is in the coordinated tactics that can be employed. A favorite tactic employed against tracking radars, for example, is to switch between jammers located on separate aircraft within the radar's beamwidth. This blinking has the effect of introducing artificial glint into the radar tracking circuits, which, if introduced at the proper rate (typically 0.1 to 10 Hz), can cause the radar to break angle track. In addition, blinking has the desirable effect of confusing radiation homing missiles that might be directed against the jammer radiations.[3]

Stand-forward jamming is an ECM tactic in which the jamming platform is located between the weapon systems and the strike vehicles and jams the radars to protect the strike vehicles. The stand-forward jammer is usually within the lethal range of defensive weapon systems for a considerable time. Therefore, only the use of relatively low-cost remotely piloted vehicles might be practical; they can assist strike aircraft or missiles in penetrating radar-defended areas by jamming, ejecting chaff, dropping expendable jammers or decoys, acting as decoys themselves, and performing other related ECM tasks.

A *self-screening jammer* (SSJ) is used to protect the carrying vehicle. This situation stresses the capability of an ECM system relating to its power, signal-processing, and ESM capabilities.

Self-protection (SP) decoy jamming is an off-board technique intended to create angle deception by causing a missile seeker to transfer angle track from the target to a decoy. Consequently, the missile guides toward the decoy and away from the target. A self-protection decoy is most likely to be used by large fighter/attack and bomber aircraft. The SP decoys are expendable or towed. Expendable decoys are ejected (or dropped) from the aircraft whereas towed decoys are tethered behind the aircraft. Expendable decoys contain miniature jamming systems that are small enough to fit into a standard chaff/flare dispenser. The decoy orients itself to the air stream by deploying low-drag aerodynamic fins sufficient to maintain stable flight. The decoy diverges from the velocity vector of the launch aircraft by natural deceleration as a result of air stream and falling due to gravity. The decoy typically starts radiating jamming signals toward the missile seeker immediately after ejection from the aircraft and continues radiation throughout its flight. Decoy ejection is typically commenced when the RWR detects incoming radar-guided missiles. Multiple decoys are sometimes dispensed at predetermined rates in order to improve the cumulative probability of aircraft survival.

A *towed decoy* is a small aerodynamically stable body that houses a miniature jammer. The decoy is deployed by reeling it out on a cable behind the aircraft to a fixed distance or offset. This offset is chosen such that even if a missile hits the decoy, the aircraft will not be damaged. The decoy can either be powered by the aircraft via the cable or be self powered. Besides providing power to the decoy, the cable can also be used as a data link to control jammer operation. Once deployed, the towed decoy can begin radiating jamming signals toward the missile seeker. When the towed decoy is no longer needed, it is either reeled in or jettisoned. The major drawback with towed decoys is that they might severely degrade aircraft maneuverability.

According to the platform, the jammer can be classified as airborne, missile-borne, naval-based, or ground-based.

A special class of missile-borne threat is the anti-radiation missile (ARM), having the objective of homing on and destroying the victim radar. The sorting and acquisition of radar signals is preliminarily made by an ESM system; afterward it cues the ARM, which continues homing on the victim radar by means of its own antenna, receiver, and signal processor. Acquisition depends on the direction of arrival (DoA), operating band, carrier frequency, pulse width, PRI, scan rate, and other parameters of the victim radar. An ARM homes on the continuous radiation from the radar sidelobes or on the flash of energy from the main beam. ARM benefits from the one-way-only radar signal attenuation. However, ARM receiver sensitivity is affected by mismatching losses; accuracy in locating the victim radar is affected by the limited dimension of the ARM antenna.

24.5 OBJECTIVES AND TAXONOMY OF ECCM TECHNIQUES

The primary objective of ECCM techniques when applied to a radar system is to allow the accomplishment of the radar intended mission while countering the effects of the enemy's ECM. In greater detail, the benefits of using ECCM techniques may be summarized as follows: (1) prevention of radar saturation, (2) enhancement of the signal-to-jamming ratio, (3) discrimination of directional interference, (4) rejection of

false targets, (5) maintenance of target tracks, (6) counteraction of ESM, and (7) radar system survivability.[3]

There are two broad classes of ECCM: electronic techniques (Section 24.6–24.9) and operational doctrines (Section 24.10). Specific electronic techniques take place in the main radar subsystems, namely, the antenna, transmitter, receiver, and signal processor. Table 24.1 shows a categorization of some ECCM techniques along with the ECM techniques that are used to counter.[5,28] Suitable blending of these ECCM techniques can be implemented in the various types of radars, as discussed in Section 24.11.

The ensuing description is limited to the major ECCM techniques; the reader should be aware that an alphabetically listed collection of 150 ECCM techniques and an encyclopedia of ECCM tactics and techniques can be found in the literature.[8,29] Many other references describe the ECCM problem, among which Slocumb and West,[5] Maksimov et al.,[21] Gros et al.,[30] and Johnson and Stoner[31] are worth noting.

TABLE 24.1 ECCM Techniques Versus ECM Technique Countered (*Reproduced with permission from Slocumb and West[5] © Artech House 2000 and G. V. Morris[28]*)

		ECM Technique Category Countered				
Radar Subsystem	ECCM Technique	Noise	False Target	Range Gate Pull Off	Velocity Gate Pull Off	Angle
Antenna related	Low or ultra-low sidelobes	×	×			
	Monopulse angle tracking					×
	Low cross-polarized response					×
	SLB (sidelobe blanking)	×	×			
	SLC (sidelobe canceler)	×				
	Electronic scan		×	×		×
	Adaptive receive polarization					×
	Cross polarization cancellation					×
Transmitter related	Low cross-polarized antenna					×
	High power	×				
	Pulse compression	×				
	Frequency diversity	×				
	Frequency agility	×	×			
	PRF jitter		×	×		
Receiver related	RGPO memory nulling			×		
	Bandwidth expansion		×		×	
	Beat frequency detector	×		×		
	Cover pulse channel processing		×			
	Home-on-jam	×				
	Leading/trailing edge track			×		
	Narrowband doppler noise detector	×	×			
	Velocity guard gates		×		×	
Signal processing related	VGPO reset		×		×	
	Signal realism		×	×	×	
	Acceleration limiting		×	×	×	
	Censored or ordered statistic CFAR	×	×			
	Doppler/range rate comparison			×	×	
	Time average CFAR	×				
	Total energy test	×				

24.6 ANTENNA-RELATED ECCM

Because the antenna represents the transducer between the radar and the environment, it is the first line of defense against jamming. The directivity of the antenna in the transmission and reception phases allows space discrimination to be used as an ECCM strategy. Techniques for space discrimination include antenna coverage and scan control, reduction of main-beam width, low sidelobes, sidelobe blanking, sidelobe cancelers, and adaptive array systems. Some of these techniques are useful during transmission, whereas others operate in the reception phase. Additionally, some are active against main-beam jammers, and others provide benefits against sidelobe jammers.

Blanking or turning off the receiver while the radar is scanning across the azimuth sector containing the jammer or reducing the scan sector covered are means to prevent the radar from looking at the jammer. Certain deception jammers depend on anticipation of the beam scan or on knowledge or measurement of the antenna scan rate. Random electronic scanning effectively prevents these deception jammers from synchronizing to the antenna scan rate, thus defeating this type of jammer. A high-gain antenna can be employed to spotlight a target and burn through the jammers. An antenna having multiple beams can also be used to allow deletion of the beam containing the jammer and still maintain detection capabilities with the remaining beams. Although they add complexity, cost, and possibly weight to the antenna, reduction of main-beam width and control of coverage and scan are valuable and worthwhile ECCM features of all radars.

If an air defense radar operates in a severe ECM environment, the detection range can be degraded because of jamming entering the sidelobes. On transmit, the energy radiated into spatial regions outside of the main beam is subject to being received by enemy RWRs or ARMs. For these reasons, low sidelobes are desirable on both receive and transmit (see Schrank,[32] Patton,[33] and Chapter 2 in Farina[34]). Sometimes the increase in main-beam width that results from low sidelobes worsens the problem of main-beam jamming; this consequence should be carefully considered in specifying the antenna radiation pattern.

Usually, specification of the sidelobes as a single number (e.g., −30 dB) means that the peak of the highest sidelobe is 30 dB below the peak of the main beam. The average, or root-mean-square (rms), sidelobe level is often more important. For example, if 10% of the radiated power is in the sidelobes, the average sidelobe level is −10 dB, where dB refers to the number of decibels by which the average sidelobe level is below the gain of an isotropic (ideal) radiator. In theory, extremely low sidelobes can be achieved with aperture illumination functions that are appropriately tapered. This leads to the well-known tradeoffs among gain, beamwidth, and sidelobe level.[35] In order to keep the beamwidth small with low sidelobes, a larger and costly (the cost could not be that large unless the radar uses an active aperture) antenna is needed. The chief problem with the low sidelobe antenna in its early days was that it had more mechanical problems because it was a waveguide array and not a reflector. Other design principles involved in low antenna sidelobes are the use of radar-absorbent material about the antenna structure, the use of a fence on ground installations, and the use of polarization screens and reflectors. This means that very low sidelobe antennas are costly in terms of size and complexity when compared with conventional antennas of similar gain and beamwidth characteristics. Second, as the design sidelobes are pushed lower and lower, a point is reached where

minor error contributions to scattered energy (random errors) or misdirected radiation (systematic errors) become significant. In practice, peak sidelobe levels as low as −30 to −35 dB (average level, −5 to −20 dB) can be readily realized with phased-array antennas that electronically scan. To obtain sidelobes at levels −45 dB down from the main beam (average level, below −20 dB), the total phase-error budget is required to be in the order of 5° rms or less. This is difficult in arrays that electronically scan: the errors induced by phase shifters, active components, and feed elements must be included in this budget. Arrays have been realized in practice that have peak sidelobes in the vicinity of the −45 dB level; however, these are generally mechanically scanned, and the low error budgets are achieved by using all-passive feed components. Phased-array developments, which do scan electronically, also foresee fairly good sidelobe performance; see the following references[36–40] for a view of relevant developments.

Two additional techniques to prevent jamming from entering through the radar's sidelobes are the so-called sidelobe blanking (SLB) and sidelobe canceler (SLC). An example of the practical effectiveness of the SLB and SLC devices is presented in the literature, where the plan position indicator (PPI) display is shown for a radar, subject to an ECM, equipped with and without the SLB and SLC systems.[31]

Other discrimination means are based on polarization. The polarization characteristics of a radar can be exploited as ECCM techniques in two ways. First, the cross-polarized pattern (i.e., the orthogonal polarization to the main plane of polarization) of a radar antenna should be kept as low as possible consistent with radar system cost. Ratios of copolarized main-beam peak gain to cross-polarized gain anywhere in the antenna pattern should be greater than 25 dB to provide protection against common cross-polarized jamming. This is thought of as an ECCM technique, but it is really no more than good antenna design. The cross-polarized jamming in this case attacks a design deficiency in the radar. The requirement for good cross-polarization design practice in a radar antenna system extends to any auxiliary ECCM antennas as well. If their cross-polarized gains are high, ECCM techniques such as SLC and SLB may not be effective against cross-polarized noise or repeater jammers.[29]

In the second use of polarization, the radar antenna system purposely receives the cross-polarization component of the radar wave in addition to the copolarized component. The two orthogonally polarized components can be used to discriminate the useful target from chaff and jammer on the basis of their different polarizations.[41] However, limited benefits (a few decibels of cancellation ratio) can be obtained at the expense of a more complex antenna system (consider, for example, a phased-array with radiating elements able to separately receive and possibly transmit the two orthogonal components of a radar wave) and of a duplication of the receiver and signal processing.

Sidelobe Blanking (SLB) System. The purpose of an SLB system is to prevent the detection of strong targets and interference pulses (as they might appear after pulse compression) entering the radar receiver via the antenna sidelobes. Thus, SLB is mainly used to eliminate interference from other pulse transmissions and deliberate pulse-like jamming. Also, SLB is effective against coherent repeater interference (CRI); here "coherent" means that the interference tries to mimic the coded waveform radiated by the radar appearing as a spike signal after pulse compression.[42–45,34] A method of achieving this is to employ an auxiliary antenna coupled to a parallel receiving channel so that two signals from a single source are available for comparison.

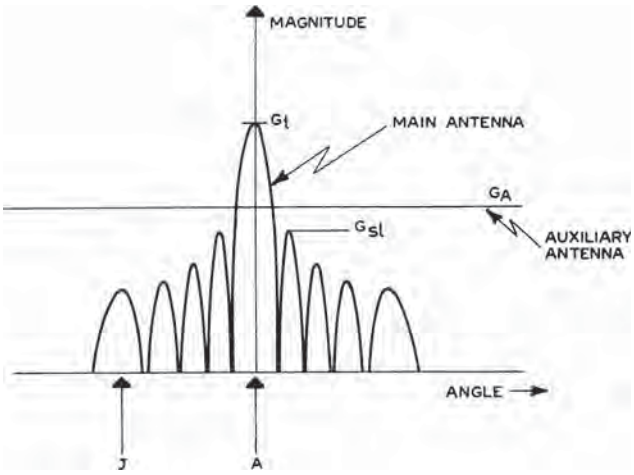


FIGURE 24.1*a* Main and auxiliary antenna patterns for the SLB (*after L. Maisel*[42] *© IEEE 1968*)

By suitable choice of the antenna gains, one may distinguish signals entering the sidelobes from those entering the main beam, and the former may be suppressed. Figure 24.1*a* illustrates the radiation pattern of the main antenna together with a low-gain auxiliary antenna. An implementation of the SLB processor is shown in Figure 24.1*b*, where the square-law-detected outputs of the two channels, ideally identical except for the antenna patterns, are compared. The comparison is made at each range bin for each pulse received and processed by the two parallel channels. Thus, the SLB decides whether or not to blank the main channel on a single-sweep basis and for each range bin. A target A in the main beam will result in a large signal in the main receiving channel and a small signal in the auxiliary receiving channel. A proper blanking logic allows this signal to pass. Targets and/or jammers J situated in the sidelobes give small main but large auxiliary signals so that these targets are suppressed by the blanking logic. It is assumed that the gain G_A of the auxiliary antenna is higher than the maximum gain G_{sl} of the sidelobes of the radar antenna.

The performance of the SLB may be analyzed by looking at the different outcomes obtained as a consequence of the pair (u, v) of the processed signals (see Figure 24.1*b*).

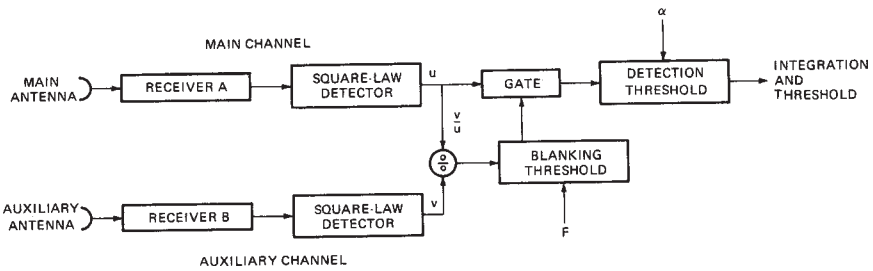


FIGURE 24.1*b* Scheme of sidelobe-blanking system (*after L. Maisel*[42] *© IEEE 1968*)

Three hypotheses have to be tested: (1) the null hypothesis H_0 corresponding to the presence of noise in the two channels, (2) the H_1 hypothesis pertaining to the target in the main beam, and (3) the H_2 hypothesis corresponding to target or interference signal in the sidelobe region. The null and H_1 hypotheses correspond to the usual decisions of "no detection" and "target detection," respectively. The blanking command is delivered when H_2 is detected.

SLB performance can be expressed in terms of the following probabilities: (i) The probability P_B of blanking a jammer in the radar sidelobes, which is the probability of associating the received signals (u, v) with H_2 when the same hypothesis is true; P_B is a function of the jammer-to-noise ratio (JNR) value, the blanking threshold F, and the gain margin $\beta = G_A/G_{sl}$ of the auxiliary antenna with respect to the radar antenna sidelobes. (ii) The probability P_{FA} of false alarm, which is the probability of associating the received signals (u, v) with the hypothesis H_1 when the true hypothesis is H_0; P_{FA} is a function of the detection threshold α normalized to the noise power level and of the blanking threshold F. (iii) The probability P_D of detecting a target in the main beam, which is the probability of associating the received signal (u, v) with H_1 when the same hypothesis is true; P_D depends, among other things, on the signal-to-noise power ratio SNR, P_{FA}, and the blanking threshold F. (iv) The probability P_{FT} of detecting a false target produced by a jammer entering through the radar sidelobes. P_{FT} is the probability of associating (u, v) with H_1 when H_2 is true; it is a function of JNR, the thresholds α and F, and the gain margin β. (v) The probability P_{TB} of blanking a target received in the main beam. This is the probability of associating (u, v) with H_2 when H_1 is the true hypothesis. P_{TB} is related to SNR, F, and the auxiliary gain $w = G_A/G_t$ normalized to the gain G_t of the main beam. To complete the list of parameters needed to describe the SLB performance, the last figure to consider is the detection loss L on the main-beam target. This can be found by comparing the SNR values required to achieve a specified P_D value for the radar system with and without the SLB. L is a function of many parameters such as P_D, P_{FA}, F, G_A, JNR, and β. A numerical evaluation of these performance parameters can be found in the literature (specifically Chapter 3 of Farina,[34] among others[42–50]).

The SLB design requires the selection of suitable values for the following parameters (Chapter 3 of Farina,[34]): (i) the gain margin β and then the gain G_A of the auxiliary antenna, (ii) the blanking threshold F, and the normalized detection threshold α. The a priori known parameters are hypothesized to be the radar sidelobe level G_{sl} and the values of SNR and JNR. The design parameters can be selected by trying to maximize the detection probability P_D while keeping at prescribed values the probabilities P_B and P_{FA} and trying to minimize P_{FT}, P_{TB}, and L. The choice of the position of the auxiliary antenna has an impact on SLB performance in presence, for instance, of multipath; to avoid its effect, the phase centers of main and auxiliary antennas should be positioned at the same height with respect to the terrain surface.

In modern radar, the blanking of sidelobe impulsive interference may be achieved by the comparison of signals, pertaining to the same cells of the range-filter map (RFM) of main beam and SLB channels. The RFM is a two-dimensional map collecting the radar echoes of all range cells (after pulse compression) and all doppler filters of a radar burst. The two RFMs are independently generated for the main and auxiliary signals and the testing of the main and auxiliary received power values is performed for all range cells and all doppler filters. This is different from a conventional SLB approach (such as the one illustrated in Figure 24.1) operating so that if interfering/repeater jammer power is detected at a particular range cell, then that range cell has to be effectively blanked. The RFM-based SLB logic greatly reduces

the risk of successfully emulating a useful target because a repeater has to appear in the same target range cell and has to emulate the same target doppler.

Sidelobe Canceler (SLC) System. The objective of the SLC is to suppress high duty cycle or even continuous noise-like interferences (NLI) (e.g., SOJ) received through the sidelobes of the radar. This is accomplished by equipping the radar with an array of auxiliary antennas used to adaptively estimate the DoA and the power of the jammers and, subsequently, to modify the receiving pattern of the radar antenna to place nulls in the jammers' directions. The SLC was invented by P. Howells and S. Applebaum.[51–52] A sample of subsequent references on SLC are also in the literature.[34,53–55]

The conceptual scheme of an SLC system is shown in Figure 24.2. The auxiliary antennas provide replicas of the jamming signals in the radar antenna sidelobes. To this end, the auxiliary patterns approximate the average sidelobe level of the radar receiving pattern. In addition, the auxiliaries are placed sufficiently close to the phase center of the radar antenna to ensure that the samples of the interference that they obtain are statistically correlated with the radar jamming signal. It is also noted that as many auxiliary antennas are needed as there are jamming signals to be suppressed. In fact, at least N auxiliary patterns properly controlled in amplitude

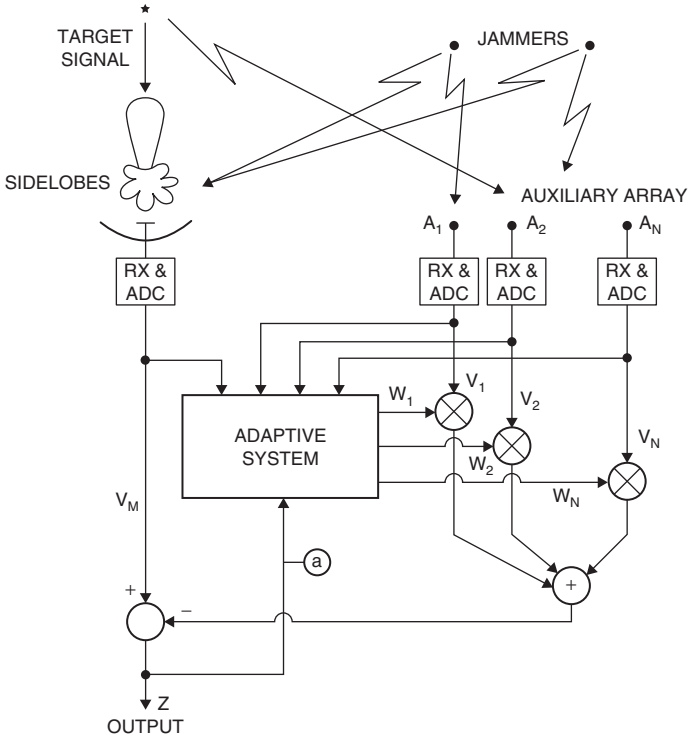


FIGURE 24.2 Principle of SLC operation (connection *a* only in the closed-loop implementation techniques); RX: receiver

and phase are needed to force to zero the main antenna receiving pattern in N given directions. The auxiliaries may be individual antennas or groups of receiving elements of a phased-array antenna.

The amplitude and phase of the signals delivered by the N auxiliaries are controlled by a set of suitable weights: denote the set with the N-dimensional vector $\mathbf{W} = (W_1, W_2, \ldots, W_N)$. The jamming signal is canceled by a linear combination of the signals from the auxiliaries and the main antenna. The problem is to find a suitable means of controlling the weights $\mathbf{W}$ of the linear combination so that the maximum possible cancellation is achieved. Owing to the stochastic nature of the jamming signals in the radar and in the auxiliary channels and to the hypothesized linear combination of signals, it is advisable to resort to the techniques of linear prediction theory for stochastic processes. Denote with V_M, the radar signal at a certain range bin and with $\mathbf{V} = (V_1, V_2, \ldots, V_N)$ the N-dimensional vector containing the set of signals, at the same range bin, from the N auxiliary antennas. It is assumed that all the signals have bandpass frequency spectra; therefore, the signals can be represented by their complex envelopes, which modulate a common carrier frequency that does not appear explicitly. The jamming signals in the channels may be regarded as samples of a stochastic process having zero mean value and a certain time autocorrelation function. For linear prediction problems, the set of samples $\mathbf{V}$ is completely described by its N-dimensional covariance matrix $\mathbf{M} = \mathrm{E}(\mathbf{V}^*\mathbf{V}^T)$, where $E(\cdot)$ denotes the statistical expectation, the asterisk $(\cdot)^*$ indicates the complex conjugate, and $\mathbf{V}^T$ is the transpose vector of $\mathbf{V}$. The statistical relationship between V_M and $\mathbf{V}$ is mathematically represented by the N-dimensional covariance vector $\mathbf{R} = E(V_M\mathbf{V}^*)$. The optimum weight vector $\hat{\mathbf{W}}$ is determined by minimizing the mean square prediction error, which equals the output residual power:

$$P_Z = E\{|Z|^2\} = E\{|V_M - \hat{\mathbf{W}}^T\mathbf{V}|^2\} \tag{24.2}$$

where Z is the system output. It is found that the following fundamental equation applies:

$$\hat{\mathbf{W}} = \mu\mathbf{M}^{-1}\mathbf{R} \tag{24.3}$$

where μ is an arbitrary constant value. The benefit of using the SLC can be measured by introducing the jammer cancellation ratio (JCR), defined as the ratio of the output noise power without and with the SLC:

$$JCR = \frac{E\{|V_M|^2\}}{E\{|V_M - \hat{\mathbf{W}}^T\mathbf{V}|^2\}} = \frac{E\{|V_M|^2\}}{E\{|V_M|^2\} - \mathbf{R}^T\mathbf{M}^{-1}\mathbf{R}^*} \tag{24.4}$$

By applying Eqs. 24.3 and 24.4 to the simple case of one auxiliary antenna and one jammer, the following results are found:

$$\hat{\mathbf{W}} = \frac{E\{V_M V_A^*\}}{E\{|V_A|^2\}} \triangleq \rho \qquad JCR = \frac{1}{1 - |\rho|^2} \tag{24.5}$$

It is noted that the optimum weight is related to the correlation coefficient ρ between the main signal, V_M, and the auxiliary signal, V_A; high values of the correlation coefficient provide high values of JCR.

The problem of implementing the optimum-weight set (Eq. 24.3) is essentially related to the real-time estimation of $\mathbf{M}$ and $\mathbf{R}$ and to the inversion of $\mathbf{M}$. Several processing

schemes have been conceived that may be classified in two main categories: (1) closed-loop techniques, in which the output residue (connection a in Figure 24.2) is fed back into the adaptive system; and (2) direct-solution methods, often referred to as *open-loop*, which operate just on the incoming signals V_M and $\mathbf{V}$. Broadly speaking, closed-loop methods are cheaper and simpler to implement than direct-solution methods; one of several practical implementations is described in Griffiths.[56] By virtue of their self-correcting nature, they do not require components that have a wide dynamic range or a high degree of linearity, and so they are well suited to analogue implementation. However, closed-loop methods suffer from the fundamental limitation that their speed of response must be restricted in order to achieve a stable and not noisy steady state. Direct-solution methods, on the other hand, do not suffer from problems of slow convergence but, in general, require components of such high accuracy and wide dynamic range that they can only be realized by digital means. Of course, closed-loop methods can also be implemented by using digital circuitry; in which case, the constraints on numerical accuracy are greatly relaxed, and the total number of arithmetic operations is much reduced by comparison with direct-solution methods. The majority of implementations has become open loop with digital technology.

Practical considerations (see Chapter 4 of Farina[34] for a detailed analysis) often limit the SLC nulling capabilities to a *JCR* of about 30 to 40 dB, but their theoretical performance is potentially much higher. Adequate cancellation of the directional interference is obtained if the receiving channels are properly matched in amplitude and phase across the radar receiving bandwidth. This condition is necessary to attribute the amplitude and phase differences measured across the channels only to the nature (power and DoA) of the impinging interference. There are several sources of mismatching; the imperfect matching of the analogue receiving channels is one of the main limitations to the interference cancellation. The effect of this mismatch on the *JCR* has been studied in the literature; see Farina[57] and references therein.

For contemporary presence of amplitude and phase mismatches, the JCR has an expression that is derived in Appendix 2.[57] A numerical application of this equation is shown in Figure 24.3; the parameter values of the study case are quoted in Farina.[57] Figure 24.3 shows the JCR contour curves versus the normalized amplitude a_n and the phase b (degrees) mismatches of the analogue receiving channels (see Farina[57] for the precise definition of these parameters).

It is seen that to have 40 dB of *JCR*, one needs to specify tight requirements for both amplitude (below 1%) and phase (below 0.7°) mismatches. This figure motivates the need to resort to equalization digital filters to compensate for the mismatches of the auxiliary channels (in their analogue part) with respect to the main channel. This subject is covered in Farina[57] and references therein. Examples of other possible limitations to cancellation are listed below[34,53,58,59]:

1. Mismatch between the main and auxiliary signals including the propagation paths, the patterns of the main and auxiliary antennas, the paths internal to the system up to the cancellation point, and the crosstalk between the channels[60–62]
2. The limited number of auxiliary channels adopted in a practical system as compared with the number of jamming signals
3. Aperture-frequency dispersion, often expressed in terms of aperture-bandwidth product[37,59,63]
4. The limited bandwidth of the majority of the schemes implementing Eq. 24.3, as compared with the wideband of a barrage jammer that can be regarded as a cluster, spread in angle, of narrowband jammers
5. Quadrature errors in synchronous (i.e., I, Q) detectors[64–66]

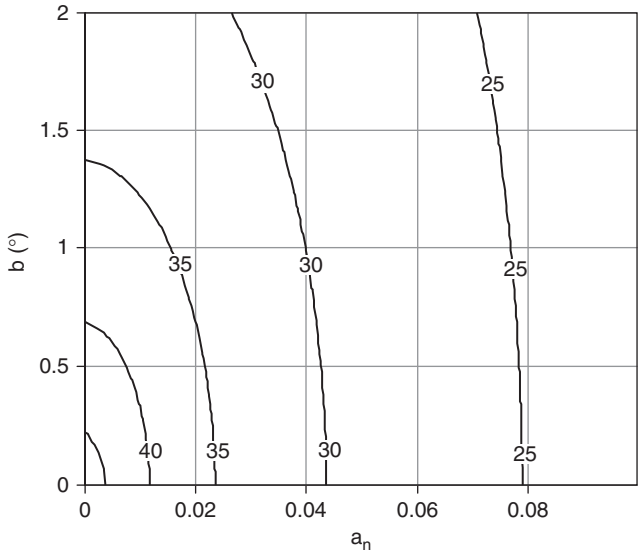


FIGURE 24.3 Contour curve of *JCR* (dB) versus the amplitude (in natural number, along the horizontal axis) and phase (in degrees, along the vertical axis) mismatches of the analogue receiving channels

6. Digital receiver channel errors such as ADC quantization, sample/hold jitter, and digital converter offset[67,68]
7. The pulse width that limits the reaction time of the adaptive system, in order to avoid the cancellation of target signal
8. The target signal in the auxiliary array that may result in nonnegligible steering of the auxiliaries toward the main-beam direction
9. Multipath delay, often expressed in terms of delay-bandwidth product[69–70]
10. The presence of clutter that, if not properly removed, may capture the adaptive system, giving rise to nulls along directions different from those of the jammers[37,71]
11. The tradeoff that has to be sought between the accuracy of weights estimation and the reaction time of the adaptive system
12. The limited number of time samples available to estimate the jammer covariance matrix; usually 3*N* sample should be available if *N* is the number of adaptive channels[101]
13. The antenna rotation rate that might produce a fast time varying power and jammer DoA[101]

Joint SLB and SLC. SLB is effective against spiky signal after pulse compression (like CRI), whereas SLC combats the continuous NLI. As previously stated, both techniques combat the interferences impinging on the main antenna sidelobes. The two techniques can be jointly used against the simultaneous presence of CRI and NLI. An approach is to cascade the SLC and SLB techniques as shown in Figure 24.4. The scheme depicts three receiving channels, each one having an antenna, a receiver, and an ADC; they provide three signals labeled, respectively, as *SLC, MAIN*, and *SLB*. The left-hand side antenna is a low-gain auxiliary performing the SLC processing in

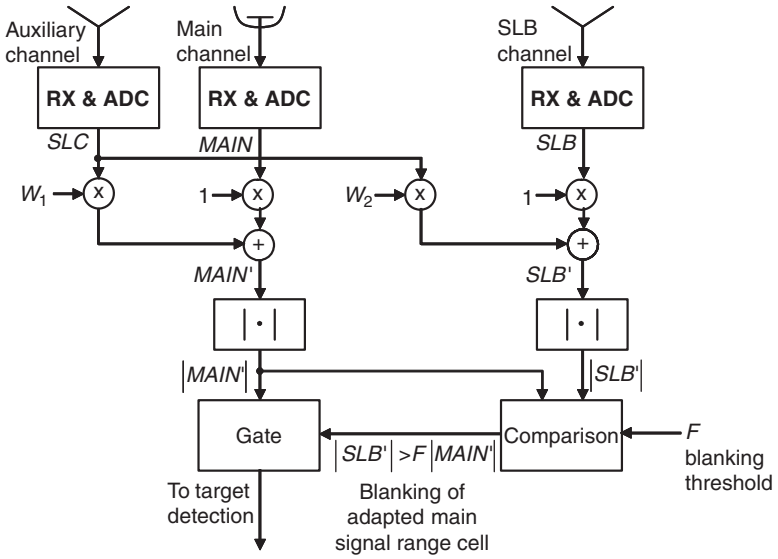


FIGURE 24.4 A processing scheme incorporating SLC and SLB devices

the main and sidelobe blanking channels. The center antenna is the high-gain radar antenna to detect targets notwithstanding impulsive and noise-like interferences. The right-hand side antenna is a low-gain auxiliary that is used for SLB processing in the main channel. The adaptive cancellation of NLI received by the main antenna is achieved by the linear combination of the *SLC* and *MAIN* signals with the adaptive weights W_1 and 1, respectively; the resulting adapted signal *MAIN*' doesn't contain the NLI. Similarly, the adaptive cancellation of NLI received by the rhs auxiliary antenna is reached by the linear combination of the *SLC* and *SLB* signals with the adaptive weights W_2 and 1, respectively; the adapted signal *SLB*' doesn't contain the NLI. Once the NLI is removed from the two channels, then the classic SLB logic can be applied against the CRI by comparing the amplitude $|MAIN'|$ of the main channel with that $|SLB'|$ of the blanking channel, which are both NLI free.[72]

Because the phase centers of the three antennas (the main and the two auxiliaries) are spaced, in general, more than 0.5 λ (where λ is the length of the radiated EM wave), the adapted patterns of the main and SLB channels fluctuate around average curves due to the presence of grating lobes.[72] Nevertheless, a reasonable gain margin is present between the pattern of the adapted SLB and the sidelobes of the adapted main antenna; thus, an adequate probability of blanking the CRI in the presence of adaptively nulled NLI should be expected. In order to improve the above gain margin and, consequently, the blanking probability of CRI, the following processing strategies are suggested[72,73]: spatial and frequency diversity.

Spatial Diversity. The rationale is to use two low-gain auxiliaries (instead of one, as shown in Figure 24.4) for the SLB; because their phase centers will be different, the grating lobes affecting the adapted patterns of the two SLB antennas will be different too. Taking the greater of the two adapted SLB signals, the gain margin between the SLB and the main antenna sidelobes will increase with a consequent improvement of the performance of the blanking logic.

Frequency Diversity. Another technique to improve the blanking performance is to resort to the diversity of the radar carrier frequency; in this case, we need just one low-gain antenna (as shown in Figure 24.4) for the SLB. The radar operates in frequency diversity mode, i.e., it radiates a burst of L pulses (T seconds apart) with slightly different carrier frequencies.[73] The grating lobes in the adapted main and the L SLB patterns will change as a function of the carrier frequency. Taking the max of the output of the L SLB signals is equivalent to a smoothing of the grating lobes. In a specific example presented in Farina and Timmoneri,[73] two carrier frequencies are used and the values of d/λ (where "d" is the inter-element distance) for the array of receiving elements are respectively 0.5 and 0.55. The blanking is separately applied on the received data at the two carrier frequencies; subsequently, the separate blanking bits are processed by a logic OR (the global blanking logic). The ensuing Figure 24.5 displays the blanking curves for the two separate carrier frequencies and for the logic OR. It is noted that the frequency diversity and the logic OR provide an improvement of the blanking probability; this is due to the different shapes of the antenna patterns at the two slightly different carrier frequencies. Figure 24.5 also presents the probability of blanking a useful target (P_{TB}) received by the main antenna beam. The probabilities are estimated via 200 independent Monte Carlo simulations. The target SNR is 20 dB; the JNR is 20 dB; the target DoA is assumed to be evenly distributed in the main-beam angular interval [–4°, 4°]; details on the numerical parameters used in the study case are in the reference.[73] It is noted that P_{TB} is negligible for $F = 0$ dB, while $P_B \geq 0.9$.

After a careful performance evaluation of the system depicted in Figure 24.4, it might be necessary to always resort to either spatial or frequency diversity to improve the SLB performance. The selection of one of the two diversity techniques depends on overall system considerations related to the impact of adding more auxiliaries and/or radiating, with the radar, proper carrier frequencies. Furthermore, if compact and high-speed processing are requested, spatial and frequency diversity techniques can be fruitfully implemented resorting to systolic schemes.

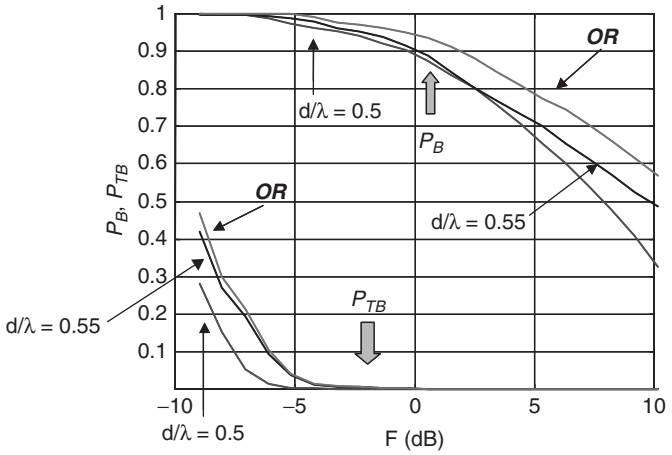


FIGURE 24.5 Blanking probability (P_B) and target blanking probability (P_{TB}) versus the blanking threshold F (in dB) for the frequency diversity scheme

Systolic Schemes for SLB and SLC. In the quest for efficient parallel processing, the systolic schemes come into the scene; their use has been described for the implementation of SLC and more general adaptive array problems in the literature[34,74,75] The rationale and the use of a systolic array that processes the signals received by SLC and main channel is reported on pages 146–156 of Farina[34] and in Farina and Timmoneri.[73] Figures 1 to 4[73] depict the use of a systolic processing scheme that incorporates the SLB and SLC. The advantage of these schemes resides in the decomposition of the complex processing for adaptive cancellation of the NLI into a network of simple processing elements that can be conveniently mapped onto a parallel processing architecture based either on Commercial off the Shelf (COTS) technology or custom Very Large Scale Integration (VLSI) devices. In the literature[76–80] it has been shown that a wide spectrum of technologies can be used like Field Programmable Gate Arrays (FPGA), Coordinate Rotation Digital Computer (CORDIC) implemented with VLSI and optical computers. Pioneering work on the use of CORDIC for adaptive nulling dates back to C. Rader at MIT-Lincoln Laboratory.[81,82] The advantage of systolic implementation is high processing speed and compact, low weight, low power consumption hardware.

Adaptive Arrays. An adaptive array (Figure 24.6) is a collection of N antennas with their own receivers (RX) and ADC, feeding a weighting and summing network, with automatic signal-dependent weight adjustment to reduce the effect of unwanted

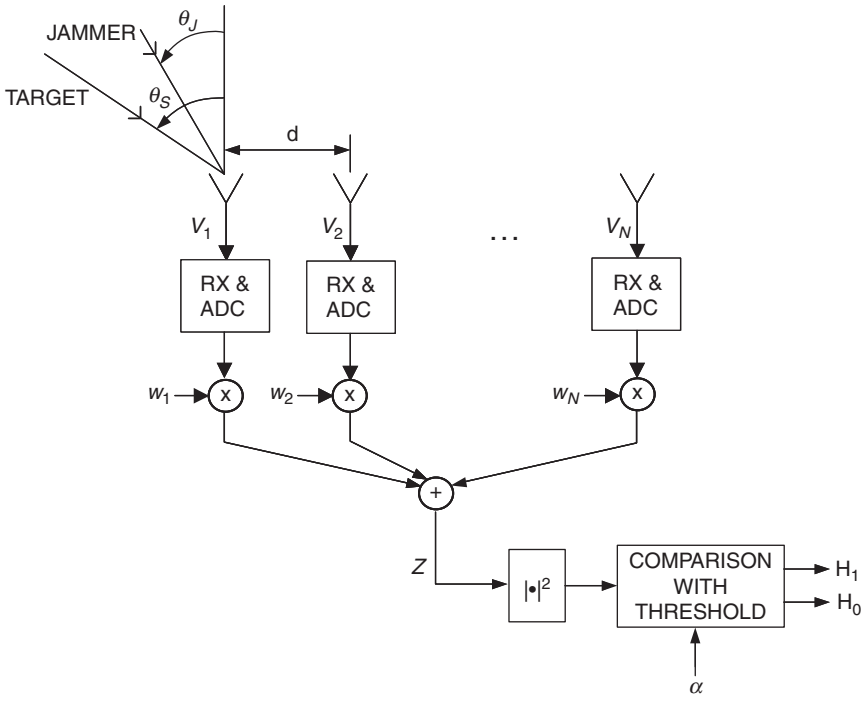


FIGURE 24.6 The adaptive array scheme

signals and/or to emphasize the desired signal or signals in the summing network output. Output signal Z is envelope-detected and compared with a suitable threshold α to detect the presence of a useful target (see Chapter 5 in Farina[34] and other sources.[53–57,83,84]) The adaptive array is a generalization of the SLC described in the preceding subsection. The basic theory of jammer cancellation and target enhancement is considered first; attention is then focused on the following topics: main-beam jammer cancellation, target DoA estimation in presence of jammer, two-dimensional adaptive processing for joint clutter and jammer cancellation, adaptivity at the subarray level, and superresolution. The implementation of the adaptive array concept is more and more related to digital beamforming[85–87] and to digital array radar (DAR) technologies.[88,89]

Jammer Cancellation and Target Signal Enhancement. Adaptive array principles have found a thorough mathematical treatment since the late 1960s[83,84]; for a brief history of adaptive arrays, see Reed[90]; for an overview of least squares adaptive processing in military applications with celebration of B. Franklin medal to B. Widrow for pioneering work on adaptive signal processing, see Etter et al.[91] The theory and application of adaptive array principles to radar is well established; for a look to popular publications see, for instance, Haykin and Steinhardt,[92] Smith,[93] and Farina et al.[94] The basic result is given by the expression of the optimum set of weights:

$$\hat{\mathbf{W}} = \mu \mathbf{M}^{-1} \mathbf{S}^{*} \tag{24.6}$$

where $\mathbf{M} = E(\mathbf{V}^{*}\mathbf{V}^{T})$ is the N-dimensional covariance matrix of the overall disturbance (noise and jammer) $\mathbf{V}$ received by the array, and $\mathbf{S}$ is the N-dimensional vector containing the expected signal samples in the array from a target along a certain direction of arrival. The similarity of Eq. 24.6 to Eq. 24.3 governing the SLC is immediately recognized.

With respect to SLC, adaptive array techniques offer the capability of enhancing the target signal while canceling the disturbance. The adaptive system allocates in an optimum fashion its degrees of freedom to the enhancement of the target signal and to the cancellation of jammer.

Several generalizations of the basic theory have been considered, including: (i) the target model $\mathbf{S}$ is not known a priori, as it is assumed in deriving Eq. 24.6; (ii) in addition to spatial filtering, doppler filtering is performed to cancel clutter and chaff; and (iii) the radar platform is moving as in shipborne, airborne, or even spaceborne applications. A relevant advancement of the adaptive array concept is space-time adaptive processing (STAP).[95–98]

STAP may be thought of as a two-dimensional adaptive filter that combines receive beamforming and doppler filtering. A basic illustration of STAP is given in Figure 1 of Ward[95] where a pictorial view of the interference environment seen by an airborne radar and the corresponding adapted two-dimensional filter response are shown. The power spectral density resulting from jammer and clutter is depicted as a function of the spatial (i.e., the sin(angle)) and the temporal (i.e., doppler) frequencies. Barrage noise jamming appears as a wall localized in angle and distributed all over doppler frequencies. The clutter echo from a single ground patch has a doppler frequency that depends on the angle between the clutter patch and the platform flight direction; clutter from all angles lies on a diagonal ridge across the space-time frequency plane. A main-beam target competes with both main-beam and sidelobe clutter as well as jamming.

The STAP generates a space-time filtering response with a main beam along the expected doppler frequency and angle of arrival of target and deep nulls along the jammer wall and the clutter ridge. To perform STAP, the radar should have an array of N antennas, each with their own receiving channel and ADC. Each channel receives M echoes from a transmitted train of M coherent pulses. Adaptivity involves the NM echoes.

The detection probability P_D for the optimum filter of Eq. 24.6 is for a constant cross section target model[84]:

$$P_D = Q\left(\sqrt{2 \bullet \mathbf{S}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{S}^*},\ \sqrt{2\ln(1/P_{\mathrm{FA}})}\right) \tag{24.7}$$

where $Q(.\,,.)$ is the Marcum Q function[‡] and P_{FA} is the prescribed probability of false alarm. It is also shown that the set of weights of Eq. 24.6 provides the maximum value of the improvement factor I_f, which is defined as follows:

$$I_f = \frac{\text{signal–to–interference plus noise power ratio at the output}}{\text{signal–to–interference plus noise power ratio at the input}} \tag{24.8}$$

The I_f value corresponding to the optimum set of weights of Eq. 24.6 is[84]

$$\mathrm{I}_f = \frac{\mathbf{S}^T\mathbf{M}^{-1}\mathbf{S}^*}{(SINR)_I} \tag{24.9}$$

The signal-to-interference plus noise power ratio $(SINR)_I$ is measured at the input of a receiving element of the array and refers to one echo pulse. The I_f represents the performance of the adaptive array: it accounts for the target signal integration and the interference cancellation. Practical applications of the equation above are, for instance, in Chapter 5 of Farina.[34] Crucial for the understanding of the adapted array pattern is the concept of eigenvalue-eigenvector decomposition of the interference covariance matrix $\mathbf{M}$: see again Chapter 5 of Farina[34] and Testa and Vannicola.[99] An important technique that mitigates the deleterious effects of the noise eigenvectors, thus continuing to maintain a prescribed level of low sidelobes in the adapted array pattern is the so-called diagonal loading.[100,101]

Adaptive arrays came about after the successful application of SLC, the application of Eq. 24.6, and of more general and powerful adaptive array concepts (e.g., GSLC: generalized SLC[34]). Clearly the efficiency of the adaptive array depends on the number of degrees of freedom (dof) and the accuracy of receiving channels (e.g., degrees of matching). There is some trade-off between accuracy and number of channels; a system with one dof is less efficient (and requires maximum accuracy) than a system with, say, four dof. An adaptive system with N dof can theoretically suppress $(N - 1)$ jammers, realistically—as a rule of thumb—$N/2$ or $N/3$. If the number of jammers is higher, the adaptive array is still useful because some jammer suppression is achieved with an accordingly reduced detection range.

---

‡ The Marcum Q function is defined as

$$Q(a,b) = \int_b^\infty x \exp\left\{-\frac{x^2 + a^2}{2}\right\} I_o(ax)dx$$

where $I_o(\cdot)$ is the modified Bessel function of order 0.

Concerning the comparison between adaptive beamforming and a very low sidelobe antenna, it is in the important case of close to main-beam jamming where adaptive beamforming is superior. On the other hand, the adaptive array allows one to obtain certain lowered sidelobes simultaneously to jamming nulling. Concerning the practical applicability of adaptive arrays some considerations follow. A number of operational radar systems are adaptive; they are described in the technical literature.[38–40,102] A modern radar with digital processing already has at least four digital channels (sum, difference in azimuth, difference in elevation, and guard). In general, the number of implemented receiving channels is mainly a matter of cost. It has been argued that radar systems with a number of adaptive dof of a few tens are already in operation in the microwave band; the number of adaptive dof may be more in the over-the-horizon (OTH) radar.

For the foreseeable future, the fully adaptive array (i.e., with adaptivity at receiving element level) has only theoretical value for antennas with a thousand elements. There are radars that are fully adaptive, but they have only a limited number of elements that can be economically handled in an adaptive array. Arrays with a large number of receiving elements need some form of processing reduction. One method of partial adaptivity is to arrange the array elements in subgroups that form the inputs of the adaptive processor. Careful selection of the subgroup elements is necessary to avoid grating lobes; this topic is discussed in a following section. Another simplification of the fully adaptive array is the deterministic spatial filtering, where a fixed reduction of the sidelobes is operated in those directions or solid angles from which the interferences are expected to come. As an example, a probable region with interferences is the horizon or part of it because jammers are mostly ground-based or at long range. The weights are computed offline, by assuming an a priori known covariance matrix **M**, and stored in a memory where a "menu" of weights is available to an operator or an automatic decision system (pp. 277–283 of Farina[34]).

Main Beam Cancellation (MBC) Systems. The objective of the MBC is to suppress high duty cycle and NLI received through the main beam of the radar. The conceptual scheme of MBC is analogous to the scheme of SLC; however, high gain beams are employed in lieu of low gain auxiliary antennas. Jamming is cancelled by a linear combination of the signals from the high gain beams and the main antenna. The weights to be applied can be computed by Eq. 24.3. The capability to cancel a certain number of main-beam interferences depends on the available number of high gain beams. A so-called four-lobed pattern can be used for main-beam interference cancellation.[103,104] The use of low-gain auxiliary antennas joined to high gain beams allows the contemporaneous cancellation of sidelobe and main-beam interferences.

Target DoA Estimation in Presence of Sidelobe and Main-Beam Interferences. Phased-array radars are required to detect, locate, and track targets in the presence of natural interference and jamming. Monopulse is the technique of choice to determine the target angular coordinates when ECM is encountered since it is much harder to deceive than a conical scan. However, the application of adaptive beamforming (to better mitigate the presence of an intense jammer) with the related distortion of sum and difference beam shapes may introduce errors in the conventional monopulse technique, in particular, if the jammer is close to the main beam[105]; thus, the conventional monopulse technique cannot be applied.[106] A Maximum Likelihood (ML) approach for target DoA estimation is considered, which generalizes the monopulse concept.[107–114,104]

The target angular coordinates—azimuth and elevation (θ, ϕ)—can be estimated by ML, also in the presence of main-beam and sidelobe jamming, by processing the data received by a set of low and high gain beams. The set of received radar echoes, $\mathbf{V} \equiv b\mathbf{S}(\theta_T, \phi_T) + \mathbf{d}$, depends on the angular coordinates of the target, (θ_T, ϕ_T), the complex target amplitude b, and white gaussian zero mean noise plus jamming disturbance $\mathbf{d}$. $\mathbf{S}$ is a vector containing the values of the patterns of high and low gain antennas in a certain direction (θ, ϕ). The data $\mathbf{V}$ are characterized by a gaussian probability density function conditioned to the target unknown parameters, i.e., $p_v(\mathbf{V}/b, \theta_T, \phi_T)$. The ML estimation of the target's unknown parameters is obtained as follows:

$$(\hat{b},\hat{\theta}_T,\hat{\phi}_T) = \arg\min_{b,\theta,\phi}\left\{\left[\mathbf{V} - b\mathbf{S}(\theta,\phi)\right]^H \mathbf{M}_d^{-1}\left[\mathbf{V} - b\mathbf{S}(\theta,\phi)\right]\right\} = \arg\min_{b,\theta,\phi}\left\{F(b,\theta,\phi)\right\} \tag{24.10}$$

where $\mathbf{M}_d$ is the disturbance covariance matrix, $\mathbf{M}_d = \sigma_n^2 \cdot [\mathbf{I} + JNR \bullet \mathbf{S}(\theta_J, \phi_J) \bullet \mathbf{S}(\theta_J, \phi_J)^H]$, depending on the angular coordinates of the jamming (θ_J, ϕ_J)[§] and on the jamming-to-noise power ratio, $JNR = P_J/\sigma_n^2$; in Eq. 24.10, $(\cdot)^H$ stands for the complex conjugate transpose operation. The amplitude b can be separately estimated by nulling the first derivative of the function to be minimized. By replacing the amplitude estimation $\hat{b}$ into the function to be minimized, the following DoA estimator is obtained:

$$(\hat{\theta}_T,\hat{\phi}_T) = \text{argmax}_{\theta,\phi}\{U(\theta,\phi)\} = \text{argmax}_{\theta,\phi}\left\{\frac{\left|\mathrm{S}^H(\theta,\phi) \bullet \mathbf{M}_d^{-1} \bullet \mathbf{V}\right|^2}{\mathrm{S}^H(\theta,\phi) \bullet \mathbf{M}_d^{-1} \bullet \mathrm{S}(\theta,\phi)}\right\} \tag{24.11}$$

It can be noted that the numerator of the functional U(θ, ϕ) is the squared adapted output ($|S^H(\theta,\phi) \bullet \mathbf{M}_d^{-1} \bullet \mathbf{V}|^2$) of a generalized array of high and low gain antenna patterns; the denominator [$\mathbf{S}^H(\theta,\phi) \bullet \mathbf{M}_d^{-1} \bullet S(\theta,\phi)$] is a normalizing term that, as we will see in a moment, plays a key role. The U function for a certain pair of angles (θ, ϕ) determines, after comparison with a suitable threshold, if a target is detected. The same U functional when scanned across a suitable set of (θ, ϕ) angle values provides, by means of Eq. 24.11, the target DoA estimate. We refer to Eq. 24.11 and its practical implementation as the *generalized monopulse technique*.

The algorithm needs the estimation of the disturbance covariance matrix $\mathbf{M}_d$, which is obtained by the radar echoes corresponding to range cells adjacent to the cell under test where a potential target is sought. The maximum of the U functional can be estimated by an exhaustive search in the range of values of interest of (θ, ϕ) or by using a fast recursive algorithm.[114] The recursion can be initialized with the angular coordinates of the main-beam pointing. By replacing the estimated disturbance covariance matrix into the U functional, a Constant False Alarm Rate (CFAR) detector is obtained.[115] Thus, the comparison of the U functional with a suitable threshold permits the target detection maintaining the prescribed CFAR. Only for the range cell in which the detection occurred, the radar signals are taken and further processed by the ML algorithm to produce the target DoA estimate.

The performance of the ML estimation algorithm of target DoA can be studied by resorting to the Cramer-Rao Lower Bound (CRLB) analysis and Monte Carlo simulations.[104,107–109,113,114] In these studies, it is shown that the shape of the U functional

[§] Here, just one jammer is considered; but the mathematical approach is easily extended to more than one jammer.

depicts the presence of the target as well as of the jamming. It has been demonstrated that Monte Carlo simulation is in close agreement with CRLB analysis. It has been found that the use of the four-lobed antenna pattern in addition to the conventional monopulse beams (sum, difference in azimuth, and elevation) can improve the estimation of the target DoA in presence of a jammer.

Joint Adaptive Jammer and Clutter Cancellation. Clutter, always present in a radar, negatively affects the performance of the adaptive jammer cancellation; therefore, means have to be adopted to effectively contrast the contemporaneous presence of clutter and jammer. When heavy clutter is present, the SLC and adaptive array will attempt to minimize the power in the adapted output without differentiating between clutter and other forms of interference. In other words, the adapted pattern will contain nulls steered in the direction of the main-beam antenna. A number of techniques may be used to avoid the problems raised by the presence of clutter. A technique, particularly suitable for low-PRF (Pulse Repetition Frequency) radar, avoids the influence of close-in clutter returns on adaptive weights by simply selecting for adaptation the clutter-free ranges at the end of each PRI. This technique does not apply to radars operating in high-PRF range-ambiguous modes with significant clutter in all range cells. If the clutter and jammer cannot be separated, either in range or doppler domains, then a two-dimensional (in doppler frequency and angle) adaptive filter might be required; this is particularly true when the statistical features of clutter and jammer are not known a priori. In fact, when either the jamming or clutter statistics cannot be estimated independently of one another, it becomes difficult to design an effective spatial adaptive filter for jamming rejection or a temporal adaptive filter for clutter mitigation since the presence of one contaminates the estimation process for the other.[116] This problem is most accentuated when the clutter-to-jamming ratio approaches unity; in which case, the cascade of a spatial and a temporal adaptive processors may perform poorly. In such situations, a joint two-dimensional adaptive filtering in doppler and angle domains represents a means to cancel the composite disturbance (i.e., the superposition of jamming and clutter) jointly rather than sequentially.[117] The performance advantages of two-dimensional adaptivity shall be traded-off with the computational cost. To reduce the computational load, different computational strategies may be devised, for example, by calculating the adapted two-dimensional weights at a rate lower than the input data and applying them to the radar snapshots at their natural rate. An efficient algorithmic procedure to extract the weights, named Inverse *QR*, is detailed in Bollini et al.[75] Other possibilities are to use modern computing technologies like the FPGA, Power PC, or high-speed optical processor to support the two-dimensional adaptive processing.[80]

Adaptivity at the Sub-Array Level. For an operational phased-array radar (PAR) with thousands of elements, it is not possible to adapt directly the signals from each radiating element. It is necessary to reduce the system complexity by using sub-arrays. A sub-array is an aggregation of antenna elementary radiators; the whole antenna can be considered as an array of these super elements. Adaptive processing can be applied at the output signals of each sub-array, thus reducing the system complexity. Provided that the sub-arrays are configured reasonably, the number of sub-arrays and the receiving channel errors (e.g, channel mismatching) determine the cancellation performance. Thus, the number of sub-arrays is a trade-off between hardware complexity, cost, and achievable performance.

It is highly desirable in PAR to have low sidelobes; this is obtained by (i) fixed weighting layer with analogue technology (i.e., at the microwave element stage) to reduce the sidelobe level everywhere; (ii) fixed weights at the digital sub-array level to reach a prescribed peak-to-sidelobe ratio (PSLR); and (iii) an adaptive weighting layer with digital technology to put nulls along the jammer DoA of high directional beams (sum, difference, cluster of high gain patterns) and low gain, possibly omnidirectional, beams (e.g., guard channel: Ω). Figure 24.7 presents a simplified scheme of a modern PAR.

Formation of Sum and Difference Patterns. Consider the problem of how to form sum and difference beams with prescribed low sidelobes in a PAR with sub-arrays. A strategy is to apply a tapering at element level (i.e., in the analogue receiving section where one attenuator is generally available per element; thus one taper function is available to achieve reasonable low sidelobes for both sum and difference beams). Subsequently a fixed digital taper, after the formation of sub-arrays, is applied with a set of weights for the sum and a different set of weights for each difference channel. This is schematically illustrated in Figure 24.8 for a uniform linear array (ULA) that generates a sum and a difference channel. The figure depicts a ULA with 24 receiving elements clustered into four not overlapping and not regular sub-arrays.[118]

The calculation of analogue taper is made by resorting to the nulling of fictitious wide angle jamming, which occupies the whole angular sector where sidelobes of sum and difference beams have to be kept low. In Farina et al.,[118] it was found that the analogue tapering is a compromise between the Taylor (which is the best taper for the sum beam) and the Bayliss (which is the best taper for the difference beam), the degree of compromise being regulated by amount of fictitious JNR selected for the sum and difference beams. A numerical example, reported in Farina et al.[118], for a ULA of $N = 24$ elements and a uniform distribution of the fictitious jammer out of the main beams of the sum and difference beams, gives a PSLR of 17.5 dB and 16.5 dB for the sum and difference beams, respectively.

The next step is to derive the fixed tapers at a digital level for sum and difference beams; a suitable technique is described in Nickel.[119,120] The rationale of the approach

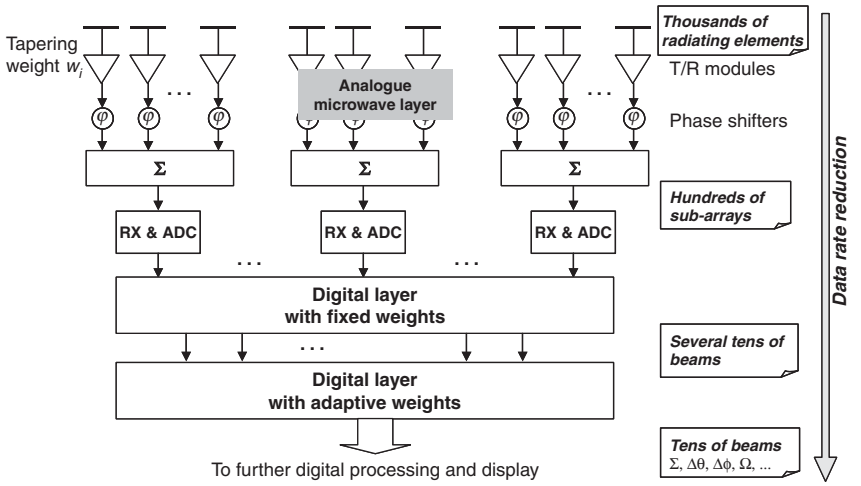


FIGURE 24.7 Scheme of a PAR

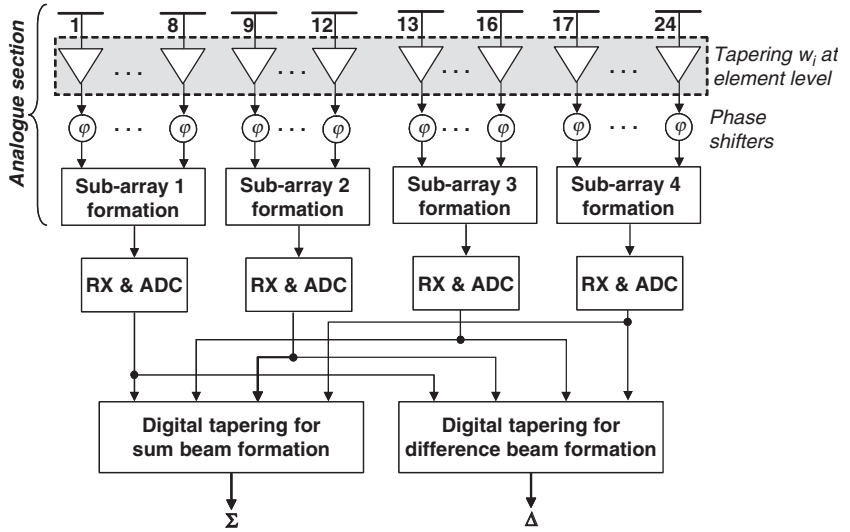


FIGURE 24.8 Example of a ULA with sub-arrays that generate sum and difference channels

is to obtain the sum beam by compensating, at the sub-array level, the analogue taper at the element level to achieve an overall taper more similar to the Taylor one; this is obtained by increasing the contribution of central sub-array weights (i.e., the sub-arrays 2 and 3 shown in Figure 24.8) with respect to the weights of the side sub-arrays 1 and 4. To obtain the difference beam, the analogue tapering at the element level is compensated, at the sub-array level, to achieve an overall taper function more similar to the Bayliss one; this is obtained by decreasing the contribution of the central sub-arrays 2 and 3.

A numerical example is reported in Farina et al.[118] with an ULA of $N = 24$ elements and $M = 4$ sub-arrays. The chosen weight is a Taylor tapering with 30 dB of PSLR. There are only 4 digital degrees of freedom; this means that marginal improvement of performance in terms of PSLR can be achieved. Nevertheless, a PSLR of 25 dB was obtained for the combination of 24 analogue weights and 4 digital weights. For the same ULA about 20 dB of PSLR was obtained for the difference channel.

Considerations Related to Sub-array Adaptivity. Tapering at the array element level produces unequal noise power at sub-array outputs because of the different number of elements in each sub-array. Adaptivity would try to equalize the noise between channels, thus negating the tapering effect.[119] The transformation $\mathbf{T}$ that encodes the sub-array architecture* should be such that $\mathbf{T}^H\mathbf{T} = \mathbf{I}$. In this way, the noise power at sub-array outputs is equal; subsequently the missing taper weights are applied digitally at sub-array outputs (weight rescaling).[119] As an example consider a linear array of 12 elements and raised cosine tapering. Figure 24.9 depicts the following. Continuous curve: pattern of array

* The sub-arrays architecture can be represented by a matrix $\mathbf{T}$; having a number M of columns equal to the number of sub-arrays and a number N of rows equal to the number of elementary radiators. The element t_{ij} of the matrix is defined either as w_i if the i-th elementary radiator belongs to the j-th sub-array or as 0 if the i-th elementary radiator does not belong to the j-th sub-array, where w_i is the tapering weight in the analogue layer of Figure 24.8.

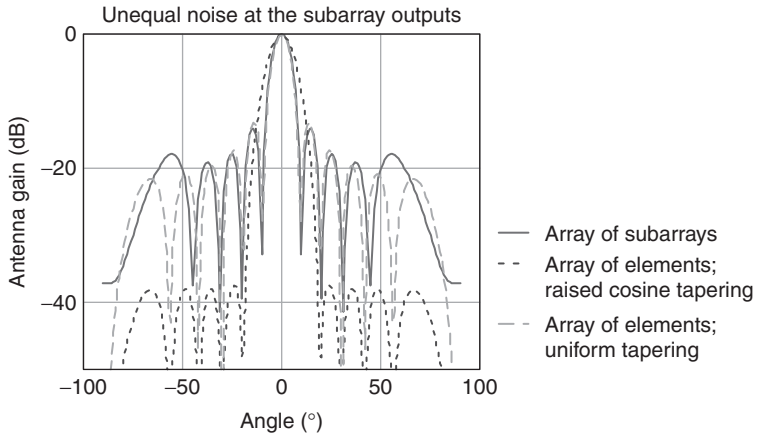


FIGURE 24.9 Examples of the antenna patterns achieved in several cases

of sub-arrays without noise normalization at the output of sub-arrays; it approximately follows the uniform tapering (dashed line). Dotted line: pattern of array of elements and of array of sub-arrays after noise normalization and weight rescaling.

The numerical example follows with Figure 24.10, which portrays the cancellation of a jammer with the DoA = –50° and a JNR of 30 dB. The continuous line refers to an unadapted pattern, tapered at the element level, whereas the dotted line pertains to the adapted pattern at the sub-array level.

Sub-arrays are, in general, chosen to be irregular in their shape and position to avoid grating lobes. If a jammer impinges on a grating lobe, the jammer will be nulled by distorting the grating lobe and, as a consequence, the array main beam (grating notch). For example, consider a ULA with $N = 12$ elements; then, form two types of not overlapping sub-array configurations both having $M = 6$ sub-arrays.

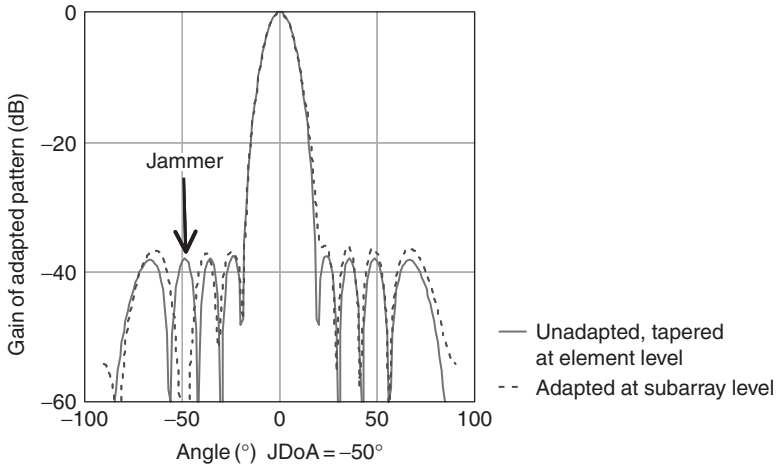


FIGURE 24.10 Jammer cancellation at sub-array level

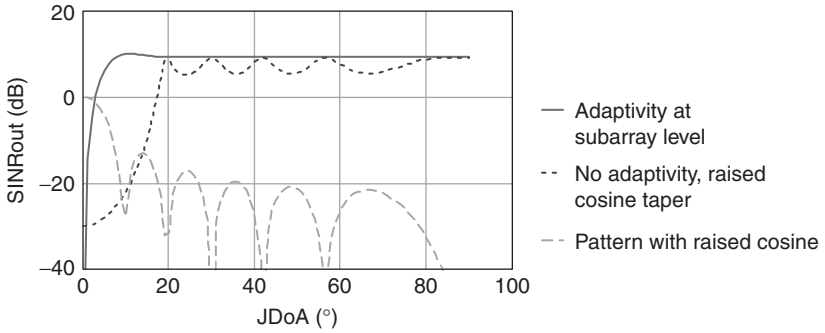


FIGURE 24.11 SINR versus the jammer DoA (JDoA)

The first configuration is regular with two elements for each sub-array. The second configuration is irregular with 2, 1, 4, 2, 1, and 2 elements, respectively. Figure 24.11 shows the SINR at the output of the array versus the jammer DoA. The target DoA is at 0°, the SNR is 0 dB, the JNR is 30 dB. Three curves summarize the system performance. The dashed line is the array pattern with raised cosine taper: this is shown for the sake of comparison with the other two curves of the SINR. The dotted line is the SINR for the quiescent (absence of adaptivity) pattern: it mimics the reciprocal of the sidelobe and main-beam pattern. The continuous line is the SINR for the adaptive irregular sub-array architecture; the maximum value of the SINR is ($10\log_{10}12$ – *tapering losses*).

Figure 24.12 depicts the SINR for the regular array configuration and absence of tapering. It is noted that when the jammer DoA is around 80°, the SINR decreases; this is because of the grating lobe. The maximum value of SINR is 10.79 dB = $10\log_{10}12$, because there are no tapering losses.

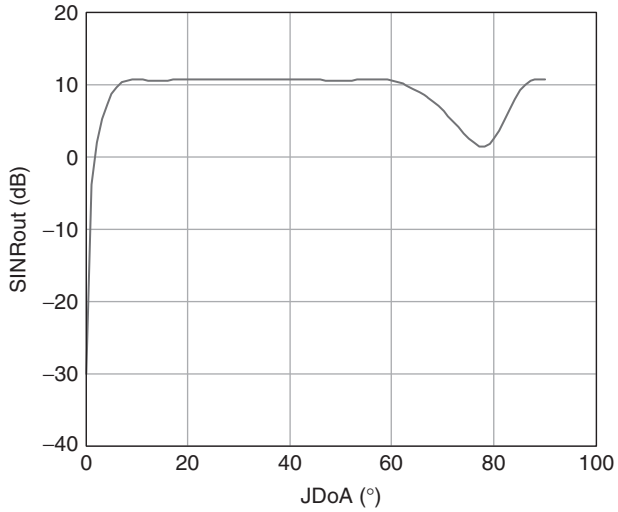


FIGURE 24.12 SINR versus the JDoA for a regular sub-array architecture and no tapering

Superresolution. The resolution of a conventional antenna is limited by the well-known Rayleigh criterion, which states that two equal-amplitude noise sources can be resolved if they are separated in angle by 0.8 λ/L, in radians, where λ is the wavelength and L is the aperture length. When the incident wave is received with a high JNR, an adaptive array antenna may—in principle—achieve a narrower *adaptive beamwidth*, giving a sharper bearing estimation of the incident wave. If accurate strobes of the jammers can be obtained, these can be exploited to form beams in the jammer directions, which are used as auxiliary channels for adaptive interference suppression.[121] The interference directions can also be used for deterministic nulling, which is of interest for main-beam nulling.[122] In addition to the interference source directions and source strengths, this technique can provide other information as to the number of sources and any cross correlations (coherence) between the sources. Such information can be used to track and catalogue the interference sources in order to properly react to them; the jammer mapping—a function running in the background—is useful to select the modes (e.g., admissible pointing directions and waveforms) of multifunction radar and for general situation awareness. Superresolution might be able to resolve multiple independent sources; due to sidelobe superposition and masking problems, superresolution might be vital for jammer mapping in case of multiple jammers. Superresolution is also of interest as an ECCM to counter cross-eye jamming in seeker head applications: see Section 12.1 of Wirth.[102]

The superresolution concept was mainly developed and analyzed by W. F. Gabriel[123] at the Naval Research Laboratory (U.S.). Different methods for bearing estimation were described by Gabriel and, subsequently, by other authors.[34,124–130] One is the maximum-entropy method (MEM) invented by J. P. Burg. It works well with a Howells-Applebaum adaptive beamformer, which has an omnidirectional receiving pattern except where jammers are present. The presence of jammers is indicated by nulls in the receiving pattern. Because nulls are always sharper than antenna lobes, jammer bearings can be obtained more accurately from the adaptive beam pattern, and superresolution is the result. The desired spatial spectrum pattern is obtained as simply the inverse of the adapted pattern. As Gabriel pointed out, there is not a true antenna pattern because there is no linear combination of the signals from an array that could produce such a peaked spatial pattern. It is simply a function computed from the reciprocal of a true adapted antenna pattern. Superresolution and adaptive antennas for jammer cancellation are intimately related. Roughly speaking, the difference is that one produces a pattern with the nulls down (adaptive antenna for jammer cancellation) and the other with the nulls up, i.e., peaks (superresolution of jammers).

One limiting factor of superresolution techniques is that they often require the received signals to obey accurate models of the array manifold. This can be violated due to propagation effects (e.g., spatial spreading and nonstationarity) as well as instrumental effects (e.g., channel mismatch). These factors also affect the performance of adaptive antennas for jammer cancellation, but model mismatches can degrade the performance of superresolution techniques more severely. The higher performance of the superresolution techniques is often obtained at the expense of requiring a stricter adherence to the assumed model; if the model is inaccurate, these techniques—which rely on its heavy exploitation—subsequently become the most sensitive and more prone to perform badly.

For efficient superresolution, an array with a reasonable number of sub-arrays is required; this may be the reason for the lack of application of this technique to practical radar systems except for experimental purposes. Superresolution based on a small

SLC configuration is not efficient, because this leads to the maximum entropy or autoregressive methods that, being nonlinear processing, have a high probability of spurious peaks.

Practical experiences indicate that the resolution limit is determined rather more by implementation and environmental factors than by the pure JNR considerations.

24.7 TRANSMITTER-RELATED ECCM

The different types of ECCM are related to the proper use and control of the power, frequency, and waveform of the radiated signal. One brute-force approach to defeat noise jamming is to increase the radar's transmitter power. This technique, when coupled with "spotlighting" the radar antenna on the target, results in an increase of the radar's detection range. Spotlighting or burnthrough modes might be effective, but a price must be paid. As the radar dwells in a particular direction, it is not looking elsewhere—where it is supposed to look. In addition, the burnthrough mode is not effective against chaff, decoys, repeaters, spoofers, and so on.

More effective is the use of complex, variable, and dissimilar transmitted signals that place a maximum burden on ESM and ECM. Different ways of operation refer to the change of the transmitted frequency in frequency-agility or frequency-diversity modes or to the use of wide instantaneous bandwidth.[131–133] *Frequency agility* usually refers to the radar's ability to change the transmitter frequency on a pulse-to-pulse or a batch-to-batch basis. The batch-to-batch approach allows doppler processing, which is not compatible with frequency agility on a pulse-to-pulse basis. In a waveform with pulse-to-pulse frequency agility, the center frequency of each transmitted pulse is moved, in either a random or a programmed schedule, between a large number of center frequencies. The frequency of the next pulse cannot generally be predicted from the frequency of the current pulse.[134] *Frequency diversity* refers to the use of several complementary radar transmissions at different frequencies, either from a single radar (e.g., a radar having stacked beams in elevation by employing different frequencies on each elevation beam[38]) or from several radars. The objective of frequency agility and diversity is to force the jammer to spread its energy over the entire agile bandwidth of the radar; this corresponds to a reduction of the jammer density and resulting ECM effectiveness.[29]

A good example of the use of the frequency domain for purpose of ECCM is the Senrad, an experimental long-range air-surveillance radar built and tested at the Naval Research Laboratory (U.S.)[135] Senrad was an example of how to build a radar so as to force the jammer to dilute its radiated energy per unit bandwidth; it includes both frequency agility and frequency diversity. This radar shows that its unusually wide bandwidth allows a reduction of the effectiveness of the noise jammer that can seriously degrade more narrowband radars.

Frequency agility, diversity, and instantaneous wideband techniques represent a form of ECCM in which the information-carrying signal is spread over as wide a frequency (or space or time) region as possible to reduce detectability by ESM and/or ARM and make jamming more difficult. This ECCM technique pertains to the realm of waveform coding.[3,136–137]

The ambiguity function (AF) is the tool to characterize waveform coding in terms of resolution, sidelobe level, and ambiguity.[137] In selecting a waveform for a given radar application, the AF should be tested against the environment in which the radar

will be expected to operate. The so-called environmental diagram depicts spectral, spatial, and amplitude characteristics of the radar environment (clutter, ECM such as chaff, intentional interferences, and—perhaps—interferences from neighboring EM apparatuses) and is used to assist the radar waveform design. An example of an environmental diagram is on p. 15 of Levanon and Mozeson[137]: on the range-doppler plane are shown the regions in which several types of clutter and high altitude chaff are expected. On the same diagram are superimposed the expected target trajectories and the AF contour of a, say, pulse-burst waveform. As the target follows a particular trajectory, the AF will move accordingly and the spurious AF peaks will slide across the clutter and chaff regions determining the intensity and features of the radar echoes.

Waveform coding includes PRF jitter and PRF stagger, which are helpful for some deception jammer but don't help against noise jammer. Waveform coding makes deception jamming or spoofing of the radar difficult, since the enemy should not know or anticipate the fine structure of the transmitted waveform; as a consequence, it gives assurance of maximum range performance against such types of jamming. Intrapulse coding to achieve pulse compression may be particularly effective in improving target detection capability by radiation of enough average radar power without exceeding peak power limitations within the radar and by improving range resolution (larger bandwidth), which, in turn, reduces chaff returns and resolves targets to a higher degree.

Some advantage can be gained by including the capability to examine the jammer signals, find holes in their transmitted spectra, and select the radar frequency with the lowest level of jamming. This approach is particularly useful against pulsed ECM, spot noise, and nonuniform barrage noise; its effectiveness depends primarily on the extent of the radar agile bandwidth and the acquisition speed and frequency tracking of an "intelligent" jammer. A technique suited to this purpose is referred to as *automatic frequency selection* (AFS).[133,138]

Another method to reduce the effect of main-beam noise jamming is to increase the transmitter frequency (as an alternative means to the use of a larger antenna) in order to narrow the antenna's beamwidth. This restricts the sector that is blanked by main-beam jamming and also provides a strobe in the direction of the jammer. Strobes from a few spatially separated radars allow the jammer to be located.

The availability of solid-state transmitter technology[36–40] allows the generation of high duty cycle waveforms, which may be of some help to realize LPI radar.

As a general remark, one of the factors preventing good ECCM is the reduction of EM spectrum allocated to radar. As discussed, operating over a wide spectral range has important advantages for ECCM, but the civilian and commercial telecommunication systems erode more and more portions of the spectrum at the expense of military ECCM capability.

24.8 RECEIVER-RELATED ECCM

Jamming signals that survive the antenna ECCM expedients can, if large enough, saturate the radar processing chain. Wide dynamic range receivers need to be used to avoid saturation.

A logarithmic (log) receiver might help against noise jamming, but it has detrimental effects against clutter when doppler processing is used. A log receiver is a device whose video output is proportional to the logarithm of the envelope of the RF input signal over a specified range. It might be useful in preventing receiver saturation in

the presence of variable intensities of jamming noise. By comparison with a linear receiver of low dynamic range, moderate jamming noise levels will normally cause the computer to saturate so that the target signal will not be detected. The main disadvantage is that a log characteristic causes spectral spreading of the received echoes. It would not be possible to maintain clutter rejection in an MTI or pulse doppler radar if the spectrum of clutter echoes were to spread into the spectral region in which target returns were expected.[21, 29]

The main message is that the dynamic range problem is important for the rejection of jammer as well as clutter, the latter always present in a radar. Thus, the recommendation is to implement, in a modern radar, a receiver with a wide linear dynamic range (e.g., 100 dB). This wide dynamic range needs to be maintained also in the ADC devices with a suitable number of bits; as a rule of thumb, each bit adds 6 dB to the dynamic range count.

Hard or soft limiters can also be used to counter jamming signals. They are nonlinear memoryless devices that cut jamming signals having wide amplitudes. The Dicke-Fix receiver counters high rates of swept-frequency CW jamming and swept spot noise jammers.[29,139] In a radar receiver, the Dicke-Fix uses a wideband intermediate-frequency (IF) amplifier and a limiter ahead of the narrow-bandwidth IF amplifier. The wideband amplifier allows a rapid recovery time from the effects of the swept jammer, and the limiter cuts the jamming signal. The narrowband target signal, after transit through the wideband amplifier and the limiter without remarkable degradation, is integrated by the narrowband filter matched to the signal. The word *Fix* in the Dicke-Fix was put there many years ago to indicate it was a "fix" for a problem that occurred at the time and was to be replaced by something better. It was usually installed with a switch to turn it off, if necessary. Today Dicke-Fix is not used in a modern radar, especially one that employs doppler processing; thus, it is no longer of interest in many radar applications.

Other special processing circuits can be used in the radar to avoid saturation, i.e., fast-time-constant (FTC) devices (perhaps of little use in modern radar), automatic gain control (AGC), and CFAR.[8,29,31] However, they cannot be said to be ECCM techniques. For example, FTC allows the detection of signals that are greater than clutter by preventing the clutter from saturating the computer. FTC does not provide subclutter visibility. AGC keeps the radar receiver operating within its dynamic range, preventing system overload and providing proper normalization so as to furnish signals of standardized amplitude to radar range, velocity, and angle processing-tracking circuits. In conclusion, these devices have a place in the radar but not as means for fighting the ECM battle.

In summary, there isn't much that has been done in the receiver to combat ECM other than to insure there is a good receiver that does its job. Today, modern phased-array multi-channel radar are going to adopt fully digital, software controlled receivers, as in the DAR case; here, the expected advantages are the wider linear dynamic range and the within-band calibration of the receivers that will support the adaptivity on several tens of channels: a distinctive advantage against directional noise jammers.

24.9 SIGNAL-PROCESSING-RELATED ECCM

Digital coherent signal processing greatly alleviates the effects of clutter and chaff.[3,140] This is motivated by the use of coherent doppler processing techniques such as fixed, adaptive MTI, or optimum pulse-doppler processing. Noncoherent devices are also

required because of the limited degree of clutter, chaff, and jammer suppression practically achieved by coherent devices, so that the cancellation residue may still be a significant source of false alarm. Among the noncoherent devices, it is worth mentioning the CFAR detector[141–145] and the pulse-width discriminator, this latter being effective against pulsed jammers. The pulse-width discrimination circuit measures the width of each received pulse. If the received pulse is not of approximately the same width as the transmitted pulse, it is rejected. A pulse-width discrimination technique can help in rejecting chaff; in fact, echo returns from chaff corridors are much wider than the transmitted pulse. However, if a target is within the chaff corridor the pulse width discriminator might also eliminate the target.

Coherent Processing. The most effective anti-chaff technique available to radar is the use of doppler filtering, which exploits the different motion characteristics of the target and the chaff.[3] The characteristics of chaff are similar to those of weather clutter, except that the chaff-scattering elements are cut to respond to a broad spectrum of radar frequencies. Weather clutter and chaff differ from ground clutter in that both the mean doppler shift and the spread are determined by wind velocity and wind shear, the latter arising from the variation of wind velocity with height. Chaff moves with the local wind, and there are ways (adaptive MTI and optimum doppler processing[†]) to make an MTI null out both moving and stationary unwanted echoes.[55,136,146] There are two basic doppler filtering techniques that are used. The first is the MTI, which employs a PRF that provides unambiguous range coverage while using a comb doppler filter whose nulls are tuned to the average radial speed of chaff.[3] The second is the pulse doppler, which can use a high PRF to provide unambiguous doppler coverage in conjunction with a doppler filter bank, allowing separation of target from the chaff.[3] A problem with chaff might be when there is a significant wind shear in the atmosphere. With wind shear, the doppler spectrum from chaff can have a wide spectral width (unless the elevation beamwidth is very narrow as might occur with tri-dimensional radar with stacked beams in elevation[38]) so that it is difficult to cancel moving chaff echoes. A pulse doppler radar has a better chance, but it has problems of its own because of the foldover of the clutter that might occupy a large range extent.

Coherent doppler processors might require relatively large amounts of pulses (e.g., more than 10), which must be transmitted at a stable frequency and PRF. A responsive jammer could measure the frequency of the first transmitted pulse and then center the jammer to spot-jam the following pulses. Also, the requirement for a stable PRF precludes the use of pulse-to-pulse jitter, which is one of the most effective techniques against deception jammers that rely on anticipating the radar transmitter pulse. Coherent doppler processors are also generally vulnerable to impulsive radio frequency interference, especially in radars with a limited number of coherent bursts on target.[147]

Another ECCM technique to be considered is pulse compression by matched filtering; it is intimately related to the waveform coding discussed in Section 24.7. Pulse compression[136,137,141] is a pulse radar technique in which long pulses are transmitted to increase the energy on the target while still retaining the target range resolution of a short pulse transmission. It is almost always used in radar for achieving

[†] The adaptive MTI estimates the mean doppler frequency of a moving clutter source and places the null of, say, a binomial MTI. Optimum doppler processing estimates the whole spectrum of clutter and shapes, by means of the inverse of the clutter covariance matrix, the cancellation filter accordingly; furthermore, with a doppler filter bank, it integrates the echo signal from moving targets. The optimum filter weights are calculated by an equation similar to Eq. 24.6 applied to the radar echoes received by a coherent pulse train transmitted by the radar.

high range resolution or reducing the peak power. Pulse compression offers some ECCM advantage that is discussed hereafter.[3,148] When the pulse compression search radar is compared, from an ESM standpoint, with a conventional search radar with the same wide pulse, the enemy receiver on a jamming platform will not know (in the general case) the pulse compression reference code and will be at a disadvantage. Compared with a radar that uses an uncompressed wide pulse, the pulse compression technique increases the radar's capability against extended signal returns like chaff and clutter. In addition, noise from a jammer does not pulse-compress. Extended clutter tends to be noise like and will not pulse-compress, which keeps down interference displayed to operators.[29] The disadvantages of pulse compression are related to the long duration of the coded pulse, which gives more time for the ECM equipment to process the pulse. In many cases, pulse compression can provide the means for easy radar jamming for the enemy ECM operator. Pulse compression is also vulnerable to cover-pulse jamming, in which the ECM pulse is returned to the radar with a high JNR such that the normal target return is covered by the jamming pulse. The width of the ECM pulse is normally wider than the radar skin return.[13] This type of deception can be counteracted by an ECCM technique such as the cover-pulse channel, where the tracking is on the ECM transmission rather than on the skin return from the target.[29]

The digital coherent implementation of the Dicke-Fix receiver concept requires the use of a coherent hard limiter that preserves the phase of the signal while keeping the amplitude at a constant value.‡ The coherent limiter is inserted upstream of the pulse compression filter in a radar that uses phase-coded signals. In reception, the jammer and target signals are cut in amplitude. The preservation of the target signal phase coding allows the integration of target energy by means of the pulse compression filter matched to the phase code. The Dicke-Fix processing scheme suffers from three limitations. The first is related to the detection loss experienced when the target does not compete with the jammer. The second disadvantage refers to the masking effect of a weak target signal sufficiently close in range (compared with the spatial extension of the code) to a strong target. Furthermore, it cannot be used in conjunction with doppler processing.

CFAR. CFAR is a technique made necessary to prevent the computer from being overloaded by false alarms, which reduce the capability of the radar to detect desired targets.[141] This processing also plays a role as ECCM; there are three motivations for this.

First, the scope of ECM techniques, in a broad sense, is to impair the target detection and tracking performance of a radar system. Detection performance is measured by the probability of detection; tracking performance is determined by the probability of detection and the probability of false alarm as well. Conventional (cell averaging) CFAR raises the threshold in the presence of noise jamming and reduces the number of targets detected. However, the targets that survive can be effectively tracked because the probability of false alarm has been maintained at sufficiently low levels. Without CFAR and appropriate threshold adjustments, perhaps no targets will be tracked due to the very large number of false peaks (detections on the jammer), making it through to saturate the tracker. Conventional CFAR is not really removing the interference: it is just "hiding" it from the radar operator. However, it is allowing the tracker to operate

‡ This is used with Barker codes, for instance, where the amplitude limitation doesn't impair the phase code.

effectively for the targets that survive, and so in this way, it can prevent the overall failure of the radar. In the limit of no targets being detected (i.e., a very powerful jammer), then it could still be argued that no tracks are better than many false tracks.

Second, not all jammers are noise jammers. Some indeed have a structure in range-doppler space, and CFAR techniques can potentially be used to lower these unwanted signals beneath the detection threshold, once again preventing the detection of false tracks, which—from a tactical perspective—can cause serious dilemmas for a radar operator.

Third, there are the adaptive CFAR detectors (AMF, or adaptive matched filter, for example[115]) that really are ECCM techniques in the sense that they enhance the probability of detection against structured interferences (in space and/or time) while maintaining the constant false alarm rate property that allows these detected targets to be effectively tracked, rather than being seduced by a high number of false detections. This type of processing—or similarly derived from the generalized likelihood ratio test (GLRT)—has been used in some practical radar.

Furthermore, any self-respecting radar system should make the operator aware of higher noise levels due to jamming even though they may not be visible on the CFAR display; the act of performing CFAR should not exclude the operator from knowing that jamming is present and that the detection threshold has been raised.

24.10 OPERATIONAL-DEPLOYMENT TECHNIQUES

To this point in the chapter, only electronic ECCM techniques have been considered. However, radar operational philosophy and deployment tactics may also have a significant effect on the radar's resistance to ECM, ESM, and ARM. This group of techniques can be subdivided into those involving the operator, the methods of operation, the radar deployment tactics, and the friend ESM in support to ECCM.[8] An operational technique against ECM is to use missiles with home on jam (HOJ) guidance to intercept noise jammers. HOJ is a means whereby a missile guidance receiver utilizes the self-screening target jamming signal to develop angular steering information so the missile can home on that target.

The role of the operator in the ECCM chain pertains to the more general topic of *human-factor* ECCM.[29] This is a generic ECCM technique that covers the ability of an air defense officer, a radar operator, a commanding officer, and/or any other air defense associated personnel to recognize the various kinds of ECM, to analyze the effect of the ECM, to decide what the appropriate ECCM should be, and/or to take the necessary ECCM action within the framework of the person's command structure. However, the human operator is less effective against a simultaneous attack of many enemy vehicles supported by a strong ECM force. An operator confronted with a large mix of ECM types and a large number of ECCM techniques is likely to do the wrong thing and/or react too slowly. In this situation, it might be proper to resort to automatically applied ECCM techniques; this is the tendency today. However, a possible concern is that this could, sometimes, hurt the ECCM capability of the radar since a well-trained operator can often figure out what is happening; but the automatic processor can only make decisions based on preprogrammed logics installed in its computer and might not recognize when something unusual (as in jamming) occurs. This is a possible adverse effect of the absence of a decision-making operator.

The operational methods include emission control (EMCON), the appropriate assignment of operating frequencies to various radars, the use of combined ECCMs to meet combined ECMs, the use of dummy transmitters to draw ECM to other frequencies, and so on. EMCON is a technique for the management of all EM radiations of a friendly system, force, or complex to obtain maximum advantages in the areas of intelligence data reception, detection, identification, navigation, missile guidance, etc., over the enemy in a given situation. EMCON permits essential operations while minimizing the disclosure of location, identification, force level, or operational intentions to enemy intelligence receptors. It includes the authorization to radiate, the control of radiation parameters such as amplitude, frequency, phase, direction, and time, the prohibition of radiation, and the scheduling of such actions for all units and equipment of a complex.[29] The on-off scheduling of the radar's operation, to include only those time intervals when surveillance is required, can reduce the probability of the radar location being found by direction finding (DF) equipment or radar homing and warning receivers. Radar blinking (using multiple radars with coordinated on-off times) can confuse an ARM seeker and guidance or a DF receiver. Decoy transmitters, radiating from antennas not located at the radar, may also be employed to confuse DF receivers and ARMs; these decoys can also operate in conjunction with the radars in a blinking mode.

Proper site selection for ground-based radars in fixed installations can provide a degree of natural signal masking to prevent, for example, detection by ground-based ESM equipment. A high degree of mobility for tactical systems allows "radiate and run" operations, which are designed to prevent the radar from being engaged by DF location techniques and associated weapons.[3] The deployment of a radar network with overlapping coverage could provide some ECCM benefits. In the netted monostatic case, the radars have different frequencies for interference reduction purposes; consequently, the ECM has to consider jamming all radars in the overlapping zone, thus reducing its efficacy. This is the kind of frequency diversity discussed in Section 24.7.

Finally, it is worth noting that friendly ESM can support ECCM action by warning of possible hostile activity, providing angular locations of hostile jammers and information characteristics of jammers. This information is helpful in the selection of a suitable ECCM action.

24.11 APPLICATION OF ECCM TECHNIQUES

This section shows the application of the previously described ECCM techniques to surveillance, tracking, phased-array, imaging, and over-the-horizon radars. The use of ECCM techniques in other types of radars such as mortar location radars, missile guidance radars, and navigation radars is considered in the literature.[3]

Surveillance Radars. The function of a surveillance radar is to search a large volume of space and locate the position of targets within the search coverage. The radar range and the azimuth-elevation coverage depend on the specific radar applications. The target reports generated by a surveillance radar are processed to form target tracks. The key features of a surveillance radar are the detection range in clear, clutter, and jamming environments, the accuracy and rate of the extracted data, and the false-alarm rate. In the ensuing discussion, the design principles, driven by the requirements forced by the threat, are mainly addressed.[3]

Detection in a clear environment is a feature of early-warning radars,[§] which look primarily for high-altitude targets at long ranges beyond the surface horizon, where the effects of clutter can be ignored. Under these conditions, a simplified analysis states that radar performance is relatively insensitive to transmitter frequency and waveform shape; in practice, the lower microwave frequencies are preferred because it is easier to obtain large antenna and high average power at lower frequencies, and rain clutter is not important. The maximum detection range on a target with a certain RCS (σ) in free space, for a surveillance radar that must uniformly search a specified volume in a given time period, depends on the product of the average transmitter power ($\bar{P}$) and the effective antenna aperture (A_r). It also depends on the inverse of the system noise temperature, but this is of little consequence since the noise temperature is not a major design issue anymore. The situation is more complex when the target to be detected is of the stealth type.[149]

Waveform design and operating frequency are relevant parameters in tactical and volume surveillance radars, which must be able to detect low-flying penetrating targets that attempt to use terrain-shielding effects to escape radar detection. In this case, the selection of waveform and frequency is made to tackle the problems of masking, multipath, chaff, clutter, and ECM.[134,138,150]

The major EW threats to a surveillance radar are (i) noise jamming, (ii) chaff, (iii) deception jamming, (iv) decoys and expendables, and (v) ARM.

Common types of jamming are main-beam noise jamming and sidelobe noise jamming. Against this threat, good radar ECCM performance is achieved by increasing the product ($\bar{P}A_r$) of average transmitter power by the effective antenna aperture. A military radar should always have 20 dB more power-aperture product than given by standard designs, yet this is seldom allowed. The request for a low sidelobe level has to be traded off with the corresponding degradation of the main-beam width; the widening of the main-beam width may make the radar more vulnerable to main-beam jamming.

The noise jammer situation is basically an energy battle between the radar and the jammer. In the main-beam noise-jamming situation, the advantage is with the jammer because the radar experiences a two-way propagation loss of its energy as contrasted with the one-way propagation loss between the jammer and the radar. With sidelobe jamming, the radar designer can reduce the jammer advantage by low sidelobe design coupled with the use of sidelobe cancellation techniques. With main-beam noise jamming, the radar can maximize the received target energy by transmitting more average power, dwelling longer on the target, or increasing the antenna gain. If the radar's data rate is fixed and a uniform angular search rate is dictated by mechanical or search strategy, then the only option for the radar is to increase its average transmitter power. The next option is to manage the data rate, thereby allowing a longer dwell time on the target (burnthrough mode) along specific spatial sectors where needed. The ability to vary the data rate in an optimal manner is one of the principal advantages of phased-array radars.[3]

Another principle of ECCM design against main-beam noise jamming is to minimize the amount of jamming energy accepted by the radar. This is accomplished by spreading the transmitted frequency range of the radar over as wide a band as possible, thus forcing the jammer into a barrage-jamming mode. This can be obtained by

[§] Of course, such radars have to see also targets at shorter ranges where clutter echoes can mask the target's echo; for this reason, all long range civil air traffic control radar employ doppler processing.

resorting to frequency agility and/or frequency diversity. Some radars incorporate an AFS device that allows the radar frequency to be tuned to that part of the spectrum containing the minimum jamming energy.[133,138]

In accordance with the search-radar equation (see Section 24.12), ECCM performance appears (explicitly) to be insensitive to frequency.* Increasing the radar frequency does not affect the signal-to-interference energy ratio within a radar resolution cell when the antenna aperture and the radar data rate are held constant. The increased frequency increases both the antenna gain and the number of radar resolution cells that must be searched by equivalent amounts; the net effect is that the target return power is increased by the same amount by which the target dwell time is decreased, thereby holding the target-to-jamming energy ratio constant. Nevertheless, in practice, the effect of main-beam noise jamming can be reduced with high radar frequency. Higher frequency radars tend to have narrower antenna beamwidths and larger operating frequency bandwidths (5 to 10 percent of radar center frequency) than lower frequency radars. Thus, main-beam jammers will blank smaller sectors of high frequency radars than of low frequency radars. In addition, main-beam jamming of a narrow beam radar tends to provide a strobe in the direction of the jammer, which can be used to triangulate and reveal the jammer location. Wider radar bandwidth, with appropriate coding, forces the jammer to spread its energy over a wider band, thereby diluting the effective jamming energy.[3]

ECCM design principles for main-beam noise jamming also apply to sidelobe noise jamming, with the addition that the sidelobe response in the direction of the jammer must be minimized. Ultralow sidelobes in the order of, say, 45 dB below the antenna's main-beam peak response are feasible by using advanced technology. Sometimes the control of sidelobe noise jamming by using ultralow-sidelobe antennas is not proper; this is true because the main-beam width might be increased two to three times. In addition, most operational radars do not use ultralow (less than −40 dB) or low (−30 to −40 dB) sidelobe antennas and have antenna sidelobes in the −20 to −30 dB region with average sidelobes of 0 to 5 dB below isotropic. SLC has the potential of reducing noise jamming through the antenna sidelobes, and it is used for this purpose in operational radars.[3]

As explained in Section 24.9, ECCM techniques against chaff are mainly those based on coherent doppler processing.[3,152] In particular, the reference[152] describes a comparison of fixed and adaptive doppler cancelers applied to chaff data recorded by a multifunctional phased-array radar operating at S band. Both cancelers process an 8 pulse coherent burst. The fixed (i.e., nonadaptive) processing is a Dolph-Chebyshev filter with 60 dB of sidelobe attenuation with respect to peak. The adaptive filter, based on the optimum doppler filtering (see Section 24.9 and the literature[55,136,146]) has the weights built around the estimation and inversion of the disturbance (chaff and noise) covariance matrix. The target detectability is evaluated against a dense chaff cloud. For the particular set of recorded measurements, it has been shown a substantially enhanced performance of the adaptive filter over the nonadaptive filter.

---

* As mentioned previously, the lower frequencies might be preferred for long-range surveillance because the usual radar equation does not include all the pertinent factors. In the jamming case, one should take account that the jamming antenna on an aircraft has a lower gain at lower frequencies so the jamming power density might be less at the lower frequencies. Also, when multipath is important, by selecting the radar frequency properly, one might reduce the jamming power received by being in a null of the jammer transmitting antenna. Chaff might not be as easy to deploy at the lower frequencies.[151] In conclusion, the lower radar frequencies might not be as vulnerable as one might think by examining the traditional radar equation.

Another class of ECCM techniques is aimed to contrast the deceptive ECM. Deception jammers have a number of specific characteristics that can be used by radars to identify their presence. The most prominent is that false-target returns must usually follow the return from the jammer-carrying target and must all lie in the same direction within a radar PRI. If the deception jammer uses a delay that is greater than a PRI period to generate an anticipatory false-target return, then pulse-to-pulse PRI jitter identifies the false-target returns. The generation of false targets in directions different from that of the jammer-carrying aircraft requires injecting pulse-jamming signals into the radar's sidelobes. Many radars employ the SLB (see Section 24.6) to defeat this type of ECM.

True-target returns tend to fluctuate from scan to scan with fixed-frequency radars and from pulse to pulse with frequency-agility radars. Transponder jammers generally send the same amplitude reply to all signals they receive above a threshold and, therefore, do not simulate actual target fluctuation responses. In addition, they usually appear wider in azimuth than real targets owing to the modulation effect of the radar scanning antenna's response on the real target. Repeater jammers can be made to simulate the actual amplitude response of real targets and, hence, are more effective over transponder-type jammers from an ECM viewpoint. An operating mode to be included in a radar to distinguish useful targets from transponder and repeater jammers is based on a doppler spectrum analysis provided enough time on target is available. Additional expensive techniques against deceptive jamming can be based on the measurement and analysis of the angular and polarization signatures of the echo signals.

The same ECCM considerations apply with decoy targets that have the general attributes of real targets and are very difficult to identify as false targets. A method sometimes employed is to test the scintillation characteristics of the detected targets to determine whether or not they follow those of real targets. Expendables that tend to be designed under stringent economic constraints often return only a steady signal to the radar. With doppler spectrum analysis, it is possible to look for returns from rotating components of the target that any form of powered target must possess. Examples are jet engine or propeller modulation returns associated with aircraft targets.

ARMs pose a serious threat to a surveillance radar. The survivability of a surveillance radar to an ARM attack relies upon waveform coding (to dilute the energy in the frequency range), the management of radiated energy in time and along the angular sectors, and the adoption of low sidelobes in transmission. These actions make it more difficult for an ARM to home on radar. When an ARM attack is detected, it may be useful to turn on spatially remote decoy transmitters to draw the ARM away from the radar site. Blinking with a network of radars achieves better results. The ARM trajectory is usually selected to attack the radar through the zenith hole region above the radar, where its detection capability is minimal. Thus, a supplemental radar that provides a high probability of detection in the zenith hole region is required. There are certain advantages in choosing a low transmitting frequency (UHF or VHF) for the supplemental radar. The RCS of the ARM becomes greater as the wavelength of the radar approaches the missile dimensions, causing a resonance effect.[3] A low-frequency radar is somewhat less vulnerable to an ARM attack owing to the difficulty of implementing a low-frequency antenna with the limited aperture available in the missile.[151] However, low-frequency radar has poor angular resolution.

Tracking Radars. Tracking radars provide good resolution and precise measurement of the kinematic parameters (position, velocity, and acceleration) of targets. The estimation, updated with measurements, and prediction of the kinematic parameters

as the time runs are the processing steps used to build up the tracks of targets. Tracks allow guidance and control of friendly forces, threat assessment, and enemy target engagement by weapons. Tracking can be accomplished in four ways: (i) The dedicated radar tracker (sometimes called a *single target tracker* and denoted STT) continuously points its antenna at a single target by sensing errors from the true target position and correcting these errors by a servo control system. Then there are two different types of radars called, in the past, *track-while-scan*. (ii) One is a limited angle scan as in some air-defense radars and aircraft landing radars, which search a limited angular sector at a rapid rate (e.g., 10 or 20 times a second). (iii) The other type of track while scan (TWS) was what is now called automatic detection and tracking (ADT). The ADT system generates tracks of more than one target by using a series of scan-to-scan target measurements taken as the antenna samples the target paths. (iv) The multifunctional phased-array radar tracks multiple targets by multiple independent beams, formed by the same array aperture, that are allotted to different targets. This subsection is limited to the design principles, driven by the threat requirements, of the dedicated radar tracker.[3,153] The ensuing subsection will refer to multifunctional phased-array radar.

Good ECCM performance is achieved by radiating as large an average transmitter power at the highest transmitter frequency practicable, coupled with as low a sidelobe level as achievable. Increasing the transmitter frequency, for a fixed antenna size, increases the antenna gain G_t, which, in turn, increases the received target power as G_t^2. For main-beam noise jamming, the received jamming power increases directly as G_t, resulting in a net increase in signal-to-jamming power by a factor proportional to the antenna gain G_t. Here, it is noted a basic difference between surveillance and tracking radar: the detection range of a tracking radar improves as the frequency is increased for a fixed-size antenna. The reason for this improvement is that the antenna gain is directly increased with frequency, thereby focusing more power on the target. This increased power is integrated for a time, which is inversely proportional to the bandwidth of the servo control loop. For a surveillance radar, this increased power is collected for a proportionally shorter time, since the radar must search more cells in the same time because of the narrower antenna beamwidth.

With sidelobe jamming, the received jamming power is proportional to the sidelobe antenna gain, (G_{sl}) resulting in a net increase in signal-to-jamming power ratio by the factor $G_t G_{sl}^{-1}$. As with surveillance radars, sidelobe noise jamming and deception can be further attenuated by the use of SLC in conjunction with SLB, as described in Section 24.6.

The use of higher transmission frequencies for tracking radars generally make them less susceptible to noise jamming than surveillance radars. In addition, tactical tracking radars may track the noise jammer in angle. Tracking a noise jammer in angle from two spatially separated radars provides enough information to locate a jammer with sufficient accuracy.

A more threatening ECM against tracking radars is DECM. These threats require considerably less energy than noise jamming (a feature particularly important on tactical aircraft, where available space is limited). Nevertheless, they are very effective in capturing and deceiving the range gate (with the RGPO technique), the velocity gate (with the VGPO technique), and the angle-tracking circuits. A primary ECCM defense against RGPO is the use of a leading-edge range tracker. The assumption is that the deception jammer needs time to react and that the leading edge of the return pulse will not be covered by the jammer. PRI jitter and frequency agility both help to ensure that the jammer will not be able to anticipate the radar pulse and lead the actual skin interval. Alternatively, the tracking radar might employ a multigate range-tracking

system to simultaneously track both the skin and false-target returns. This approach utilizes the fact that both the jamming signals and the target return come from the same angular direction, so that the radar's angle-tracking circuits are always locked onto the real target.[3]

The methodology of introducing VGPO into the radar's tracking circuits is analogous to the method used with RGPO. The frequency shift is initially programmed so that the repeated signal is within the passband of the doppler filter containing the target return. This is needed to capture the doppler filter containing the target, through the radar's AGC action. The repeater jammer signal is then further shifted in frequency to the maximum expected doppler frequency of the radar. The repeated signal is then switched off, forcing the victim radar to reacquire the target.[3] Coherent tracking radars can check the radial velocity derived from doppler measurements with that derived from differentiated range data. Anomalous differences provide a warning of the probable presence of a deception jammer. When RGPO and VGPO operate simultaneously, the best defense is the contemporary tracking of true and false targets in both range and doppler dimensions. The use of multimode (high, low, and medium PRF) radars can also be an effective ECCM measure helping to counter range-gate and velocity-gate stealers by switching radar modes.

Angle-gate stealing is particularly effective against conical-scanning or sequential-lobing tracking radars. It is for this reason that such trackers cannot be used in military applications. The fundamental problem with these radars is that angle tracking is accomplished by demodulating the amplitude modulation imposed on the target return pulses over a complete scanning or lobing cycle. To jam this type of radar effectively, the radar's angle-tracking-error-sensing circuits must be captured with a false amplitude-modulated signal, at the scanning or lobing rate, which is significantly out of phase with that from the target return. When the conical-scan or lobing modulation is imposed on both the transmitter and the receiver beams, it is relatively simple for a jammer to synthesize the appropriate jamming signal by inverting and repeating the transmitter modulation (inverse-gain repeater).[154] This can be partially overcome by a conical scan-on-receiver only (COSRO) system, where the tracking radar radiates a nonscanning transmitting beam but receives with a conical-scan beam. The jammer then has no knowledge of the phase of the conically scanned receiving beam and must adopt a trial-and-error method of scanning the jamming modulation until a noticeable reaction occurs in the tracking radar beam. (This jamming technique is called jog *detection*.[13]) A sequential lobing-on-receive only (LORO) system conceals the lobing rate from a potential jammer.[3] Conical scan and sequential lobing are going to be replaced by the monopulse technique; thus, COSRO and LORO are becoming obsolete.

Monopulse tracking is inherently insensitive to angle deceptive jamming from a single point source. This is a result of the monopulse angle-error-sensing mechanism that forms an error proportional to the angle between the target and the antenna's boresight on each return pulse. This is accomplished by comparing signals received simultaneously in two or more antenna beams, as distinguished from techniques such as lobe switching or conical scanning, in which angle information requires multiple pulses. Effective monopulse jamming techniques generally attempt to exploit the monopulse radar's susceptibility to target glint or multipath signals.[13]

One jamming approach, known as *cross-eye*, used against monopulse radars generates artificial glint into the monopulse tracking loop.[13] The inventors of the cross-eye technique are B. Lewis (NRL, USA) and D. Howard; see their patent[155] originally filed the 1958. Cross-eye is basically a two-source interferometer whose antennas usually are

mounted on the aircraft's wingtip as far apart as possible. The signals received in each wingtip antenna are repeated in the opposite wingtip antenna, except for a 180° phase shift, which is inserted in one line to direct an interferometric null toward the victim radar. In effect, this creates an apparent change of target direction as viewed from the radar. A large repeater gain is required to generate a high jammer-to-signal ratio; otherwise, the skin echo will overwhelm the jamming signals in the interferometer pattern nulls. The maximum effectiveness of the technique implies a considerable delay (on the order of 100 ns) in the repeated signal, owing to the transmission line and amplifier between the receiver and transmitter antennas. Thus, leading-edge or multigate range tracking should be an effective ECCM technique against cross-eye jamming.[3,13]

Terrain-bounce jamming or terrain scattered interference (TSI) or hot clutter is another monopulse jamming technique that is used against semiactive missile seekers and airborne tracking radars. With this technique, the jammer aircraft illuminates the Earth's surface in front of and below it, so that the semiactive missile homes on the illuminated ground spot and not on the jammer aircraft. The uncertainty of the terrain scattering parameters and the possible depolarizing effects of surface reflection are some of the problems associated with this technique.[3]

The TSI against airborne radar and the corresponding mitigation techniques are described in detail in the literature[156–158] TSI is a significant problem to military airborne radar; in fact, an often weak target signal in the main beam has to compete with jammer that propagates not only via direct-path but also via multipath from the underlying terrain. Mitigation techniques have been focused on estimating the direct jammer signal, estimating the linear system created by the multipath, and removing an estimate of the reflected jammer signal from the main received radar signal[158]; this is also allowed by using reference beams pointed at hot clutter.[157] Adaptive cancellation techniques have to be able to account for the doppler induced by relative motion between airborne radar and jammer platforms and the jammer signal nonstationarity that is produced from such a bistatic geometry. TSI mitigation for over-the-horizon (OTH) radar is described in Abramovich et al.[159]

Monopulse radars that use parabolic reflector antennas are susceptible to jamming through cross-polarization lobes generated by the reflector surface.[3,13] This occurs because the angle-error-sensing discriminator has an inverse slope for a cross-polarized signal, which causes the angle-tracking servo to have positive feedback instead of the negative feedback required for tracking. Monopulse estimates with planar array antennas usually have a high resistance to cross-polarization jamming (see Section 11.5 of Wirth[102]). With array antennas—in contrast to reflector antennas—all the single antenna elements have the same polarization-dependent pattern. This is multiplied with the array factor and also applies for the sum and difference patterns. The resultant form of the beam pattern will thus be independent of polarization. Therefore, the monopulse operation will also not be disturbed.[102]

Phased-Array Radars. In this subsection, we illustrate, by a numerical example, the role played by the scheduler in a multifunctional PAR to combat ECM. To this end, we resort to a benchmark study described in the literature, which defines typical ECM threats, operational scenarios, and phased-array performance mainly in terms of target tracking under ECM. The simulation benchmark[160] includes two types of ECM, namely SOJ and RGPO. The SOJ, mounted on an aircraft, transmits broadband noise toward the radar. The SOJ flies an oval (race course) holding pattern in a clockwise direction at an altitude of 3050 m and a speed of 168 m/s; it is approximately

150 km from the radar. The two circular turns are performed at an acceleration of 1.5 g. The transmitted SOJ noise impacts the radar with power γ_0 not exceeding eight times the receiver noise power. Thus, a SOJ will not completely hide a target, and it can be defeated with a higher energy waveform. In RGPO, the target under track repeats with delay and amplification the radar pulse so as to pull the radar range gate off the target. The time delay is controlled so the false target is separated from the true one with either linear or quadratic motion. For the linear case, the range of the false target R_k^{ft} is related to the range of the true target R_k^{t} via

$$R_k^{\text{ft}} = R_k^t + v_{\text{po}}(t_k - t_0) \tag{24.12}$$

where v_{po} is the pull-off rate, t_k is the time at which the target is being observed, and t_0 is the initial reference time of the RGPO false target. Alternatively, for the quadratic case

$$R_k^{\text{ft}} = R_k^t + \frac{1}{2}a_{\text{po}}(t_k - t_0)^2 \tag{24.13}$$

where a_{po} is the pull-off acceleration.

Radar Scheduling. The scheduling and the tracking functions closely cooperate; both interact to update, with current measurements, the target's state vector, and make the predictions necessary to point the radar beam at the target the next time it is observed, select the type of waveform to radiate, and select the threshold to apply for target detection. A conceptual scheme showing the interaction of scheduling and tracking is shown in Figure 24.13, where r_k, b_k, e_k are the range, bearing, and elevation measurements at t_k; SNR_k is the observed SNR at t_k; t_{k+1} is the commanded time for the next target observation; $r_{k+1|k}$, $b_{k+1|k}$, $e_{k+1|k}$ are the predicted range, bearing, and elevation for beam pointing control at t_{k+1}; W_{k+1} is the waveform selection at t_{k+1}; β_{k+1} is the detection threshold for the dwell set at t_{k+1}; and $\mathbf{X}_{k|k}$, $\mathbf{P}_{k|k}$ are the target filtered state estimate and covariance matrix at t_k given all the radar measurements up to t_k.

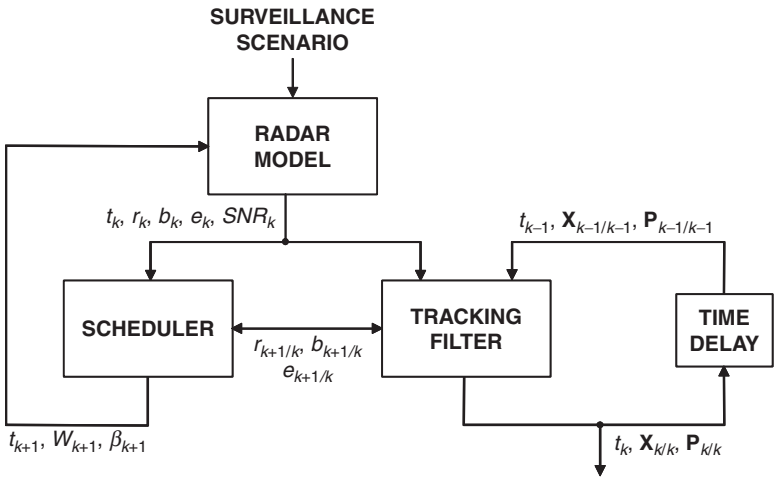


FIGURE 24.13 Interaction of radar scheduler and tracking filter

The scheme is built around two closed loops: (i) the loop encompassing the radar model, the tracking filter, and the scheduler, and (ii) the tracking filter loop. The scenario is observed by the radar at time t_k; the radar model provides the measurements r_k, b_k, e_k and SNR_k. The tracking filter updates the previous target state estimate $\mathbf{X}_{k-1|k-1}$, and its covariance matrix $\mathbf{P}_{k-1|k-1}$ at time t_k providing the new estimates $\mathbf{X}_{k|k}$ and $\mathbf{P}_{k|k}$ and the values $r_{k+1|k}$, $b_{k+1|k}$ and $e_{k+1|k}$ at the next time instant t_{k+1}. The scheduler provides the waveform to radiate W_{k+1} and the threshold β_{k+1} to apply for target detection at t_{k+1}.

Selection of the Sampling Period. The sampling period is chosen among a finite number of possible different values based on kinematic considerations on the target (estimated speed) as well as on whether missed detections have occurred. If there is no measurement to be associated to the target, the sampling period is set equal to $T_s = 0.1$ s and the waveform of highest energy is selected, so as to possibly avoid a second missed detection due to the possibly low target-RCS. Conversely, the sampling period is selected as follows:

- $T_s = 0.5$ s for targets with estimated speed greater than 400 m/s
- $T_s = 2$ s for targets with estimated speed between 100 and 400 m/s
- $T_s = 3$ s for targets with estimated speed less than 100 m/s

Even though the target may accelerate or maneuver, for the sake of simplicity, the sampling period is selected only on the basis of the target estimated speed.

Selection of the Detection Threshold. The presence of a jamming signal can increase the number of false alarms and wrong plot-track associations up to an unacceptable level, thus increasing significantly, the probability of loosing a target under track. It is, therefore, important that the radar receiver be equipped with a CFAR. Since the false alarm probability is related to the detection threshold, the latter should be adapted online based on the intensity of disturbances.

Selection of the Waveform. The benchmark[161] includes 8 waveforms, indexed by i and characterized by a different pulse width $\tau_e(i)$, so that the waveform can be selected in order to provide a SNR greater than the detection threshold and thus maintain an assigned probability of target detection. This can be accomplished by first estimating the average target RCS_k[†] at time t_k, and then computing for each waveform i, the predicted $SNR_k(i)$, and finally selecting the waveform index i such that the corresponding $SNR_k(i)$ is just greater than the desired detection threshold plus a given tolerance.[161]

ECCM: A-SOJ and A-RGPO. Hereafter, specific *anti-SOJ* (A-SOJ) and *anti-RGPO* (A-RGPO) techniques will be described.

A-SOJ is based on estimating the jammer position and power level and then using such estimates to adapt the radar detection threshold online.

- *Jammer state estimation*. Whenever the radar operates in the passive mode, i.e., without emitting pulses, bearing b_k^j and elevation e_k^j of the jammer as well as the relative standard deviations σ_k^{jb} and σ_k^{je}, and the jammer-to-noise ratio ρ_k^j (in the following

[†] The RCS is certainly a fluctuating quantity versus time; it also depends upon the target aspect angle. However, if enough time-on-target is available, the RCS estimate can be sufficiently accurate.

expressed in dB) are measured. This allows the tracking filter to estimate the jammer state made up of four state components: the two angular positions (bearing and elevation) and the relative angular speeds. The jammer track is initialized by using the first two measurements provided by the radar.

- *Jammer power level estimation*. An estimate of the power level can be obtained by the first-order linear filter initialized from $\gamma_0(t_1^j) = 1$ for a suitable filter coefficient $\alpha_j \in (0,1)$:

$$\gamma_o\left(t_k^j\right) = \alpha_j \cdot \gamma_o\left(t_{k-1}^j\right) + (1-\alpha_j)\cdot 10^{\rho_k^j/10} \tag{24.14}$$

- *Adaptation of the detection threshold*. For a given detection threshold β (in dB), the probability of false alarm turns out to be:

$$P_{\text{fa}} = \exp\left(-\frac{10^{\beta/10}}{\gamma_0 G_{\text{stc}}(R)\left(\Sigma_k^j\right)^2 + 1}\right) \tag{24.15}$$

Hence, the detection threshold can be selected at each time instant, t_k, in the following way:

$$\beta_k = \max\left\{9.64,\ 10\log_{10}\left[-\gamma_0\left(t_{k-1}^j\right)G_{\text{stc}}(r_{k/k})\left(\Sigma_k^j\right)^2 + 1\right]\ln P_{\text{fa}}\right\} \tag{24.16}$$

where $\gamma_0(t_{k-1}^j)$ is the most recent available estimate of the jammer power level; the value $\beta_k = 9.64$ dB is the one which allows, in the absence of jammers, the desired false alarm probability $P_{\text{fa}} = 10^{-4}$; $r_{k|k}$ is the filtered estimate of the target range; $G_{\text{stc}}(.)$ is the sensitivity time control gain; and Σ_k^j is the normalized antenna gain for the received signal computed in the radar active mode.[161]

Because the phased-array radar considered here is a multifunctional one, it has also a tracking mode that might be affected by the RGPO; for this reason, an A-RGPO is considered an ECCM technique. Whenever RGPO is active, two high-amplitude signals are received from the radar: the true target's echo and an RGPO-induced signal. Since the time at which the target under track activates RGPO is unknown to the tracking algorithm, the latter must first recognize that RGPO is active and then implement an appropriate A-RGPO technique. In order to establish whether RGPO is active, the following test can be adopted. Let N be the number of measurements exceeding the detection threshold of more than 3 dB. Then if $N < 2$, it is decided that RGPO is not active and no A-RGPO action is undertaken; otherwise, if $N = 2$, the A-RGPO ECCMs described below are applied. Notice that the test also aims at discriminating the type of ECM being active, i.e., SOJ or RGPO. In fact, whenever the noise jammer is in the antenna main beam many false measurements with high jammer-to-noise ratio are induced; in this case, it turns out that $N > 2$ and RGPO is declared inactive. Whenever the jammer is no longer in the target's line-of-sight, it may happen that multiple measurements exceed the detection threshold, but the condition that the excess is greater than 3 dB will be very unlikely to be fulfilled in practice. Once it has been established that RGPO is active, several devices can be adopted in order to prevent the loss of the target under track.

1. A first approach consists of maintaining two tracks until the RGPO is deactivated.
2. A second approach consists of penalizing, in the data association, the measurements whose range is greater than the average range of measurements with SNR higher than the detection threshold.[161]

3. A third, more drastic approach, consists of discarding the measurement with higher range among the two measurements that have exceeded the detection threshold of more than 3 dB.

It is important to guarantee, under RGPO, a very high SNR for the target. In fact, it might happen that the signal produced by the false target overcomes the detection threshold whereas the one from the true target does not, thus causing an association error with possible serious consequences in the target's tracking. Hence, whenever RGPO is active, a high energy waveform must be selected. A further precaution is the following: if there are missed detections for at least two out of the last three scans, an immediate revisit with sampling interval $T_s = 0.1$ s in search dwell mode is performed. In search mode, the range gate is 10 km instead of 1.5 km, so in this way, it is possible to obtain a new target measurement for updating the tracking filter and thus avoid the target's loss.

Simulation Results. Monte Carlo simulation experiments using the benchmark[160] have been carried out in order to assess the benefits of the above described ECCMs. More specifically, the adaptation of the detection threshold has been used as A-SOJ whereas the technique based on discarding the measurements with higher range has been adopted as A-RGPO. Three types of targets (numbered 1, 5, and 6) have been considered: target 1 represents a cargo aircraft while targets 5 and 6 represent fighter/attack aircrafts with a much higher degree of maneuverability. For each experiment, the following results are displayed: number of lost targets (over 50 Monte Carlo trials), T_s (radar sampling time), T_{ave} (average fraction of time required by the radar for target tracking), P_M (average power), position error, and velocity error. Table 24.2 shows simulation results for the Interacting Multiple Model (IMM) tracking[162] algorithm in the absence of ECMs. Tables 24.3 and 24.4 report the results in the presence of SOJ, without and respectively with A-SOJ. Similarly Tables 24.5 and 24.6 report the results in the presence of RGPO, without and respectively with A-RGPO. The examination of these tables reveals that the presence of ECMs considerably deteriorates the tracking performance if no appropriate ECCMs are undertaken. Conversely, the adoption of the above described A-SOJ and A-RGPO techniques allows the performance that would be attained in the absence of the corresponding ECMs to be restored.

TABLE 24.2 Simulation Results Without ECM

TARGET NO.	LOST TARGETS	T_s (s)	T_{ave} (s)	P_M (W)	POS ERR (m)	VEL ERR (m/s)
1	0	1.958	$0.5106\ 10^{-3}$	5.7985	116.8	65.26
5	1	0.6772	$1.477\ 10^{-3}$	68.898	95.39	61.29
6	1	1.112	$0.899\ 10^{-3}$	10.774	82.94	58.43

TABLE 24.3 Simulation Results With SOJ and Without A-SOJ

TARGET NO.	LOST TARGETS	T_s (s)	T_{ave} (s)	P_M (W)	POS ERR (m)	VEL ERR (m/s)
1	34	1.919	$0.521\ 10^{-3}$	6.6179	127.5	71.09
5	15	0.6923	$1.444\ 10^{-3}$	68.411	103	66.78
6	50					

TABLE 24.4 Simulation Results With SOJ and A-SOJ

TARGET NO.	LOST TARGETS	T_s (s)	T_{ave} (s)	P_M (W)	POS ERR (m)	VEL ERR (m/s)
1	1	1.944	$0.5144\ 10^{-3}$	6.6179	127.5	71.09
5	1	0.6888	$1.452\ 10^{-3}$	68.411	103	66.78
6	4	1.118	$0.8944\ 10^{-3}$	15.11	80.49	59.59

TABLE 24.5 Simulation Results With RGPO and Without A-RGPO

TARGET NO.	LOST TARGETS	T_s (s)	T_{ave} (s)	P_M (W)	POS ERR (m)	VEL ERR (m/s)
1	48	1.963	$0.5095\ 10^{-3}$	5.044	120.5	66.6
5	50					
6	50					

TABLE 24.6 Simulation Results With RGPO and A-RGPO

TARGET NO.	LOST TARGETS	T_s (s)	T_{ave} (s)	P_M (W)	POS ERR (m)	VEL ERR (m/s)
1	0	1.889	$0.5295\ 10^{-3}$	6.6179	127.5	71.09
5	1	0.7045	$1.419\ 10^{-3}$	68.411	103	66.78
6	0	1.156	$0.8651\ 10^{-3}$	15.586	124.9	80.26

Imaging Radar. There are two types of imaging radar that will be discussed: synthetic aperture radar (SAR) and inverse SAR (ISAR).

SAR. SAR allows us to have a high-resolution mapping of the EM backscatter from an observed scene. More precisely, the radar data is obtained in polar coordinates, i.e., slant range and azimuth, while a two-dimensional image in the rectangular coordinates (x, y) is provided. High resolution in slant-range is obtained by transmitting a coded waveform, with a large value of the time-bandwidth product, and coherently processing—in a filter matched to the waveform—the echo signals. High resolution along the transversal direction is achieved by forming a synthetic aperture. This requires (i) to put the radar onboard a moving platform, e.g., an aircraft or a satellite; (ii) to record the EM signals from each scatterer that is illuminated by the moving antenna beam in successive instants of time, and (iii) to coherently combine the signals—via a suitable azimuthal matched filter—thus focusing the sliding antenna pattern in a narrower synthetic beam. Radiometric resolution, another key parameter, is related to the capability of SAR of distinguishing different objects in the scene on the basis of their EM reflectivity. Radiometric resolution determines how fine a sensor can distinguish between objects with similar EM reflection properties. It is a parameter of great importance, especially for those applications oriented to extended target exploitation like polarimetry and classification. Thus, the radiometric resolution should be optimized mainly for good extended target interpretation, accounting for all kind of back-scatterers. Multilook processing is commonly used in SAR image

formation in order to reduce the speckle noise. Traditional digital multilook processing consists of an incoherent addition of independent images (looks) of the same scene. The looks can be obtained by partitioning the available signal bandwidth (range and/or azimuth) and processing each look independently. The final image is produced by adding the looks incoherently, pixel by pixel. The direct trade-off between geometric and radiometric resolution must be considered when choosing the number of looks for processing. One-look processing means a fully coherent use of the bandwidth (best geometric resolution), and in this case, the speckle noise will obey an exponential distribution where the standard deviation is equal to the mean value in the intensity image (multiplicative characteristic). For multilook processing, the geometric resolution will degrade as the number of looks increases and the speckle statistics of the intensity image obey a gamma distribution, where the standard deviation decreases with the square root of the number of independent looks.[163]

SAR images are useful for surveillance and reconnaissance applications. However, jamming could make SAR images unusable. The use of ECCM is, therefore, essential to reducing the vulnerability of SAR to jammers. The susceptibility to intercept signals from SAR and vulnerability to jamming are described in Goj.[164] A simulated noise jamming produces stripes on the SAR images that demonstrate the effectiveness of jamming against targets like strong point scatterers such as electric power-line towers, as well as low-reflectivity agricultural patterns and desert land. The references[165,166] discuss the significant vulnerability to ECM of spaceborne SAR during maritime reconnaissance missions. In 1978, typical imagery from Seasat SAR showed several features that made the SAR a powerful maritime surveillance sensor. The ships and the wakes produced by the ship motion were imaged. The ship image (a blob) appeared displaced from its wake due to the doppler shift caused by the ship's motion relative to the spacecraft. However, the ship's position at the time of imaging and its course could be determined from the wake. The ship speed can be calculated too by the displacement of the ship from its wake. All this information is obtainable only if enough SNR is available for the identification of these image features either by a human operator or an automatic processor. There exists, therefore, a potential vulnerability to SAR in the maritime surveillance application if a high level of background noise causes degradations of the SAR image to an extent where ship and wake can no longer be identified in the image. In the literature[165,166] some critical aspects, in terms of jammer receiver sensitivity and transmitted power, for spot-noise jamming are considered and the system requirements are derived to determine the feasibility and practicality of such jammer. Results of a computer simulation of an engagement between SAR and representative jamming system are given to enable the effectiveness of ECM to be assessed.

The threats to a SAR are barrage jamming, spot jamming, random pulse jamming, and repeater jamming. Repeater/deception jamming is a major threat because it might not be recognizable, whereas the others are, at least in principle, recognizable. The impact of each threat and the possible countermeasures are described in the remaining part of this section.

- *Barrage jammer.* The disturbance noise extends over the entire swath of the SAR image, and it shows generally a uniform intensity. The radar image of barrage jammer noise will exhibit speckle, i.e., a brightness variation from one resolution cell to another. In addition, because a large number of noise samples are added noncoherently, the multiple looks of jammer noise tend to smooth out the intensity variation from pixel to pixel, just as in the case of thermal noise.

- *Spot jammer*. It also covers the entire swath and with uniform intensity disturbing noise as for the barrage jammer; however, its image will differ from the barrage noise because the Fourier transform of narrower band jammer noise will result in speckle size in the range dimension that is larger than that of thermal noise or clutter. The processed cross-range dimension is again equal to that of clutter or thermal noise. Spot jammer noise will appear to be stretched in range.
- *Random pulse jamming*. The jammer pulses may also be transmitted at random intervals, so that such noise pulses can appear in any part of the range swath. When observed over a sufficient number of samples, the noise pulses will occupy all parts of the range swath in one sample or another. The azimuth processor forms the sum of the noise power from all samples within one synthetic aperture length. That sum will be equal to the total noise power in the aperture, which is proportional to the average jammer noise power. Also, in this case, the speckle dimension will appear stretched in range, just as in the spot jammer case. However, the random pulse jammer speckle will exhibit more pronounced brightness variations than that from spot or barrage jamming, because fewer noise samples are added noncoherently, thereby reducing the smoothing effect of multiple looks.
- *Repeater jamming*.[167] The enemy may utilize the transmitting radar to send out a signal within the band of the SAR to confuse the SAR system receiver. The jamming signal causes the SAR to receive and process erroneous information that results in severe degradations in the SAR images and/or formation of the image of nonexistent targets. A deception jamming could be composed of manipulated replicas of the transmitted radar signals via DRFM. In Hyberg[168] the possibility of preventing SAR mapping through coherent DRFM jamming has been investigated. A software model has been developed and verified in several flight trials in the case of a ground-based DRFM jammer.

  ECCM techniques for SAR can be divided into (i) antenna-based techniques (low sidelobes, adaptive arrays) and (ii) transmitter/receiver/processing–based techniques (frequency agility, pulse coding).

  - *Low sidelobes*. SAR antennas with low sidelobes reduce the level of jamming power received and, in addition, reduce the probability of being intercepted by ECM stations (in the sidelobe region).[169] In relation to low sidelobes, the following comment is in order. In a conventional radar, the effect of low sidelobes is clear, but there is a difference in SAR because the beamwidth is much wider than in other radar applications. In principle, the finer the resolution the smaller the SAR physical antenna and the wider is its beamwidth. Thus, jamming in the main beam is more likely in a SAR than in other radars because of the wide main beam. To get false targets into the SAR, it would have to come from the main beam, so low sidelobes might not be much involved in deception jammer. Likewise, main-beam jamming may be more of a threat to SAR than sidelobe jamming.
  - *Adaptive arrays*. The references[170–172] deal with the rejection of a barrage noise jamming using an adaptive spatial nulling. Equipping the SAR system with an antenna partitioned into several sub-apertures connected to parallel channels (i.e., multichannel SAR) allows spatial adaptive processing to suppress the interfering signal. In Farina and Lombardo,[170] the performance of such a technique is evaluated in terms of SAR impulse response, detection performance of point target, and radiometric resolution of an extended scene. In Ender[171] a SAR image, taken with an experimental four-channel SAR jammed by a small 1 watt

noise jammer leading to a JNR in the raw data of about 30 dB when the jammer passes the center of the main beam, is depicted. The de-jammed image by adaptive spatial suppression is also shown demonstrating the good performance of adaptive spatial cancellation. The reference[171] provides a comprehensive study of anti-jamming spatial adaptive techniques including also the space/slow-time anti-jamming filter with suitable image reconstruction algorithm. Results indicate that the slow-time STAP provides superior interference cancellation than spatial-only filtering. SAR usually involves wideband processing, requiring for adaptive nulling techniques, peculiar algorithms. Efficient broadband jammer nulling has to be countered with space/fast-time (i.e., range cell) processing.[172] The expected number of spatial dof is not higher; we have only to add the dof in time. The adaptive beamforming algorithms have to be implemented into the SAR processing which is always space-time processing (typically post-doppler). Rosenberg and Gray[173] tackle the problem of mitigating the effects of an airborne broadband jammer present in the main beam of a SAR. In addition to this, multipath reflections from the ground, known as hot-clutter, will add a nonstationary interference component to the image. The authors show the image degradation from hot-clutter, the limited restoration that multi-channel spatial imaging and slow-time STAP can provide, and how fast-time STAP can improve the final image quality.

- *Frequency agility*. SAR processing needs phase coherence for obtaining the synthetic aperture, thus frequency agility has to be used with care. Frequency changing during a synthetic aperture length time results in a change of focal length (different coefficient of the quadratic phase term) of the phase history of the illuminated targets that degrades the cross-range resolution. SAR operating in burst mode can change its central frequency from one look to another, without any degradation in image quality. Given the efficiency of simple broadband jamming and modern ESM, we need to conclude that frequency agility is not of great help in SAR ECCM.
- *Pulse coding*.[167] An effective ECCM against a DRFM repeat jammer is to change the radar transmitted pulse code from one PRI to another. The radar maintains the same carrier and bandwidth; however, the pulses are coded to be approximately orthogonal to each other (i.e., their cross-correlation is approximately equal to zero). Such a radar is less susceptible to a DRFM repeater because (i) the jammer cannot adapt easily since the radar signal is varying in the PRI domain, and (ii) the signal transmitted by a DRFM repeater jammer at a given PRI (i.e., the radar signal that is used by the SAR at the previous PRI) is approximately orthogonal to the radar signal that the SAR is utilizing at the current PRI, and thus, a matched filtering with the current PRI radar signal would weaken the DRFM repeater jammer signal. In Soumekh[167] a novel method is outlined that combines the above mentioned pulse diversity radar signaling with a new coherent two-dimensional processing of the measured data to effectively suppress a DRFM repeater jammer.

ISAR. The inverse SAR is a method of reconstructing a high-resolution two-dimensional EM intensity image of moving targets (e.g., ships, aircraft) in the range and cross-range (doppler) domains. ISAR imaging is important in military applications such as target recognition and classification (since it can usually recognize the class of target) that can also be used to cue weapon systems. The need for coherent countering of these imaging sensors is a high priority for EW. The references[174,175] present the

design of a pipelined all-digital image synthesizer capable of generating false-target images from a series of intercepted ISAR chirp pulses, thus providing RF imaging decoy capability. The image synthesizer modulates the phase samples from a phase sampling DRFM that stores intercepted ISAR pulses. The image synthesizer must also synthesize the temporal lengthening and amplitude modulation caused by the many reflective surfaces of a target and must generate a realistic doppler profile for each surface. The position of a false target image in range can be controlled by delaying in time the read-out samples going to the image synthesizer. The range-doppler image of a ship with 32 range bins is synthesized as an example in Pace et al.[175] ECCM techniques to defeat this type of jamming signals are similar to those proposed for SAR.

Over-the-Horizon Radar. An important defense-related role of high frequency (HF) over-the-horizon (OTH) radar is to provide a capability for early warning detection and tracking of air and ship targets. By using the ionosphere as a propagation medium, *skywave* OTH radars can operate at very long distances to achieve detection and tracking at ranges of 500–3000 km. On the other hand, *surface-wave* OTH radars exploit vertically polarized HF signals (3–30 MHz) and the conductive properties of sea water to detect targets at ranges limited to about 250 km. This upper limit generally applies to large ships and frequencies in the lower HF band.[176,177]

ECM to OTH Radar. For both skywave and surface-wave OTH radar systems, the ionosphere also propagates unwanted interference signals to the radar site, particularly at night when the ionosphere is prone to propagating radio frequency interference (RFI) sources from very long distances. RFI can arise from unintentional and intentional anthropogenic emitters in the user-congested HF band, as well as jamming sources. Jamming sources may be located on the target platform itself (self-screening) and received by the main antenna beam or radiate from a separate location (stand-off) and be received mainly through the antenna beampattern sidelobes. The jamming signal may be incoherent with the radar waveform and operate in a "spot" or "barrage" fashion to raise the noise floor in both range and doppler search spaces to potentially impair detection performance, or it can be coherent with the radar waveform, as in the case of deception jamming, which may generate false targets and potentially impair the tracking system from following the true target.

Impact of Ionosphere. An important aspect that distinguishes OTH radar from line-of-sight systems is the impact of the ionospheric propagation medium on the characteristics of the received interference. The ionosphere is stratified with different reflecting layers, so a single interference source is often received as a number of multipath components with different DoAs, both in elevation (due to the different heights of reflection points) and in azimuth (due to layer-dependent ionospheric tilts or gradients). In addition to multipath, each interference component is subjected to temporal and spatial distortions caused by the dynamic behavior of electron density irregularities present in the individual reflecting layers.[178] This physical phenomenon is known not only to deform the interference wavefronts relative to the anticipated plane wavefront, but also to induce a significant level of spatial nonstationarity on the various interference components over time intervals commensurate with the coherent processing interval of OTH radars (in the order of a few to tens of seconds).[179,180]

Relevance to Interfering Signals. Sources of interference within the radar coverage (e.g., on an airborne platform) can potentially screen the platform in range and impair the detection of other targets with similar azimuth but possibly at different ranges.

Such sources can expect good propagation to the radar receiver because the choice of operating frequency is usually optimized for the coverage area. In the case of stand-off interference sources, which are located arbitrarily with respect to the surveillance region (e.g., a ground-based emitter), propagation conditions will generally be sub-optimum. However, such sources may have greater power and antenna gain at their disposal, allowing the signals to reach the radar receiver with appreciable strength, sometimes after propagation via highly disturbed and nonstationary ionospheric paths (as commonly occurs in the equatorial and polar regions). Under normal circumstances, OTH radars seek to find relatively clear frequency channels in the user-congested HF spectrum, so the presence of interference from other manmade sources is effectively diminished by suitable frequency selection. When jamming is present, the radar may need to operate with much higher than usual levels of interference that can degrade performance. For this reason, protection in the form of ECCM techniques becomes necessary.

ECCM Techniques. Electronic protection for OTH radar antenna arrays can be provided in the form of adaptive signal processing in space and time. The stochastic constraints adaptive beamforming and STAP methods[181–184] were developed specifically for the HF environment to address the rejection of nonstationary interference while protecting the clutter doppler spectrum properties. A method for time-varying spatial adaptive processing (TV-SAP)[185] that addresses the same problem was found to be more attractive for practical implementation due to the much lower computational cost in real-time applications, as well as greater robustness in protecting sub-clutter visibility after doppler processing. The problem of reducing false alarms caused by strong sidelobe targets and spatially structured (non-gaussian distributed) RFI was treated in Fabrizio et al.,[186] where the advantages of adaptive subspace detectors relative to conventional approaches were shown. STAP techniques with temporal degrees of freedom spaced at the PRI (i.e., slow-time) have been proposed in Farina et al.[187] to jointly cancel RFI and clutter when both are of similar strength but neither can be isolated for estimation, whereas an alternative low-dimension STAP formulation with temporal taps spaced at the range cell interval (i.e., fast-time) has been proposed in Fabrizio et al.,[188] to jointly cancel sidelobe and main-beam RFI that exhibits correlation in the range dimension. The STAP methods used in OTH radar are very similar conceptually to those adopted for airborne radar, especially the former tap architecture.[187] The chief difference is that in benign conditions (free of significant co-channel interference[‡]), STAP is not indicated for OTH radar because the sidelobe clutter does not typically mask doppler shifted targets any more than the main-beam clutter.[§] A possible exception to this is shipborne HF surface-wave radar,* although such systems have been proposed, they have not yet demonstrated their practical utility.

[‡] Co-channel interference for OTH radar refers mainly to other transmissions in the HF spectrum that either fully or partially overlap the radar bandwidth.

[§] The main beam and sidelobe clutter received by airborne radars can have quite different doppler shifts due to the movement of the platform with respect to the ground resulting in the angle-doppler coupling of the clutter. However in OTH radar, main beam and sidelobe clutter from a single ionospheric mode typically have similar doppler spectrum characteristics because the radar is stationary. This means that sidelobe clutter appears at roughly the same doppler shift as main beam clutter and doppler filtering can be used effectively for detecting targets usually without special need to reject the sidelobe clutter spatially.

* Obviously the situation may change in shipborne HF surface wave radar because the platform is moving with respect to the sea surface and hence a conceptually similar situation arises for the clutter as encountered in airborne radar.

24.12 ECCM AND ECM EFFICACY

There is a need for a quantitative measurement of the efficacy of one or more ECCM electronic techniques when a radar equipped with these devices is subject to an ECM threat. One performance measure generally used for an unjammed search radar is the detection range of a certain target against a system noise background; this situation is referred to as *detection in clear environment*. When the radar is jammed, it is of interest to calculate the degradation of the detection range with respect to self-screening, standoff, and escort jammers. These calculations apply to both search and tracking radars. For tracking radars, it is also worthwhile to consider the degradation of measurement accuracy and resolution. The benefits of using ECCM techniques such as frequency agility, coherent doppler processing, very low sidelobe antennas, and SLC can be easily assessed at a first approximation by properly modifying the parameters involved in the radar equation. If, for instance, an SLC is adopted against an SOJ, its net effect is to reduce jamming power by the amount of jammer cancellation ratio that the SLC can offer.

The prediction of radar range is difficult because of the many factors that are hard to represent with models of the required accuracy. The factors involve the target to be detected (target returns of an unknown statistical nature), the natural environment in which the target is embedded (e.g., clutter returns, unintentional interference, uncontrollable environmental refraction, and absorption), the random nature of the interference, and the radar itself (system noise temperature, signal distortions, etc.). Nevertheless, radar range prediction made under average conditions provides a preliminary and useful indication of performance under ECM threat and ECCM design effectiveness that produces baseline values prior to simulation and operational tests. A classical book presents accurate detection range equations in a variety of practical situations.[189] In the second part of this section, a review of software tools available for the prediction of range equation in jamming and chaff conditions is given.

Of course, the radar equation is a simplification in assessing ECM-ECCM interactions; a measure of ECCM effectiveness should involve the whole weapon system in which the radar operates. The measure of effectiveness should be expressed in terms of the number of attackers destroyed or the probability of radar survival. References in the literature attempt to assess the ECCM efficacy.[190–194]

Simulation is another means to assess the ECCM benefits in radar and weapon systems.[193] An advantage of this approach resides in the capability to artificially generate different types of threats and to look at the radar[160,161] and weapon system reactions. However, the simulation of such a complex system is a difficult, time-consuming task that sometimes involves the use of ad-hoc programming languages suitable for simulation.

Simulation of a complex system on a digital computer is a technique used for the analysis, design, and testing of a system whose behavior cannot be easily evaluated by means of analysis or computation. The procedure essentially consists of reproducing the algorithms of a suitable model of the examined system by means of computer programs. Proper inputs to the model, corresponding to the most relevant operational conditions for the real system, can be generated by the same computer programs. The outputs obtained are compared with some reference values (expected or theoretical) to assess system performance. When random inputs are provided, a number of statistically independent trials are performed to achieve a significant sample of the output values from which reliable statistics can be estimated.

The accuracy and detail of the model may vary from a coarse functional description of the system to a very accurate one, according to the purpose of the simulation and the required accuracy of the results. However, it is desirable to limit the complexity of the simulation tools in order to have manageable programs, giving results that are easily interpreted. The accuracy in representing each system function depends upon its relevance with respect to system performance. When a very complex system is to be simulated, it is generally preferred to resort to several programs of limited complexity in lieu of a single bulky simulation. This approach corresponds to partitioning the whole system into subsystems separately modeled in detail. From each partial simulation, a limited number of relevant features are extracted and employed to build a simplified model of the overall system.

Simulation is particularly important to account for the adaptive nature (e.g., CFAR, adaptive beamforming, automatic radar management, adaptive tracking, adaptive clutter cancellation) of modern radar systems.[195] In this case, traditional static measures such as detection range against a given target will no longer adequately define the capabilities of radar systems. Measures of radar dynamic characteristics, such as the susceptibility to processor overload or the time to adapt in changing conditions, are more important. Modeling and simulations to evaluate the radar response to standardized changing scenarios represent an attractive technical solution.[195]

Simulation is always of value; however, the effectiveness of ECM and ECCM is ultimately done, when possible, with tests of real EW capabilities against real radar systems under real-world conditions. This is especially important for radar equipped with adaptive techniques since they might not be always fully modeled in a simulation as they are in the real-world environment in which they must operate.

The Radar Equation in Jamming and Chaff Conditions. An example of radar range performance under noise jamming is reported on pp. 14–19 of Farina,[34] where the important role played by a radar with low sidelobe antennas is also noted. Today the use of computer programs for predicting radar performance under jamming, clutter, and chaff, and in the presence of various refined propagation models is well established: there are programs developed in house by individual radar companies[196] or available on the market.[197]

The Radar Work Station (RWS) is an example of a developed in-house program[196]; RWS originates from the modeling and simulation activities carried out for prediction of radar performance in several scenarios. One main objective of RWS is to provide the radar analyst or system designer with a friendly but comprehensive toolkit for prediction of radar performance based on well recognized, flexible, and documented mathematical models. A broad range of radar types (bi-dimensional, multi-beam three-dimensional, phased-array), composite clutter, ECM and propagation scenarios, and a target's kinematics and RCS features are covered. Input and output data can be saved, loaded, and exported to other similar applications or for general use (i.e., MS Office tools for data analysis). A second purpose is to provide a handy and reliable tool for technicians and engineers performing system setup at the site or acceptance tests by means of field trials, by providing not only the software tools and models but also, where required, a database of prediction results, and allowing simple parametric excursions thereof, without the need to consult a bulky reference documentation. In brief, the most valuable outcomes that can be obtained with the RWS are radar range calculation, radar elevation coverage diagrams in clear, ECM and multi-propagation both for coherent and noncoherent radars; range and velocity responses in complex

scenarios (multiple clutter sources, user defined trajectory) in terms of signal-to-disturbance power ratio and detection probability; and radar range and height accuracy calculation, radar resolution evaluations employing suitable data extracting logics. The RWS suite consists of the following main modules: C\C++ and Fortran (to calculate special functions like Gamma and Bessel K) coded libraries; a standard library of Windows APIs (Application Programming Interfaces) to draw the coverage diagrams; a template library to implement the matrix algebra; an application based on MS Office Excel to code the Blake Chart†; a set of Visual Basic tools to evaluate some aspects of radar performance (e.g., ADC jitter, atmospheric loss, tapering loss, etc.); and an unformatted archive of radar data, as simple ASCII files, pertinent to performance, environment, trajectories, terrain height, and waveform. User-friendly interfaces run on low-cost platforms (PC) and popular environments (Win98, WinNT, Win2000, Windows XP, Vista) for users and developers.

In the RWS, the chaff volume clutter model is characterized in terms of volume extent, EM reflectivity, and doppler spectrum. The signal-to-noise plus chaff ratio is determined on the basis of the chaff location in the space, the antenna receiving pattern, and the radiated radar waveform. The radar equation can be applied, and the signal processing scheme can be emulated to determine the amount of chaff mitigation. A barrage noise jammer is modeled in terms of effective radiated power (ERP) and frequency band of operation. The signal-to-noise plus jammer ratio is determined on the basis of the JDoA and the antenna receiving patterns; the radar equation is then applied; and suitable ECCM signal processing schemes can be emulated to determine to which extent the jammer is attenuated.

Computer Aided Radar Performance Evaluation Tool (CARPET) is an example of available software on the market. In the CARPET 1.0 manual,[197] the equations for calculating the contributions from chaff (volume clutter) in the signal-to-interference ratio are described on pp. 59 and 60, and the equations for calculating the contribution from noise jamming (barrage or responsive) are described on p. 61. CARPET is programmed in C++ and has a Windows XP-compatible user-friendly graphical interface.

ACRONYM LIST

ADC	Analogue-to-Digital Converter
ADT	Automatic Detection and Tracking
AF	Ambiguity Function
AFS	Automatic Frequency Selection
AGC	Automatic Gain Control
AMF	Adaptive Matched Filter
API	Application Programming Interface
A-RGPO	Anti Range Gate Pull Off
ARM	Anti Radiation Missile
A-SOJ	Anti-Stand Off Jammer

† Actually, it is a generalized Blake chart (that improves on the original Blake chart), which includes details such as antenna patterns, processing, system losses, etc., in a suitable electronic format.

BM	Ballistic Missile
CARPET	Computer Aided Radar Performance Evaluation Tool
CFAR	Constant False Alarm Rate
CORDIC	COordinate Rotation Digital Computer
COSRO	Conical Scan-on-Receiver Only
COTS	Commercial Off The Shelf
CRI	Coherent Repeater Interference
CRLB	Cramer-Rao Lower Bound
CUT	Cell Under Test
CW	Continuous Wave
DAC	Digital-to-Analogue Converter
DAR	Digital Array Radar
DECM	Deceptive ECM
DF	Direction Finding
DoA	Direction of Arrival
Dof	Degree of freedom
DRFM	Digital Radio Frequency Memory
DToA	Difference Time of Arrival
EM	Electromagnetic
EA	Electronic Attack
ECCM	Electronic Counter-Counter Measure
ECM	Electronic Counter Measure
ELINT	ELectronic INTelligence
EMCON	EMission CONtrol
EP	Electronic Protection
ERP	Effective Radiated Power
ES	Electronic Support
ESM	Electronic warfare Support Measure
EW	Electronic Warfare
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Arrays
FTC	Fast Time Constant
GA	Genetic Algorithm
GSLC	Generalized Side Lobe Canceler
HOJ	Home On Jam
HF	High Frequency
IF	Intermediate Frequency
IMM	Interacting Multiple Model
ISAR	Inverse Synthetic Aperture Radar
JCR	Jammer Cancellation Ratio
JDoA	Jammer Direction of Arrival
JNR	Jammer-to-Noise Ratio
LORO	Lobing-On-Receive Only
LPI	Low Probability of Intercept
MBC	Main Beam Canceler
MEM	Maximum Entropy Method
ML	Maximum Likelihood
MTD	Moving Target Detector
MTI	Moving Target Indicator

NLI	Noise Like Interference
OTH	Over The Horizon
PAR	Phased-Array Radar
PDR	Phase Difference Rate
PDW	Pulse Description Word
Penaids	Penetration Aid Decoy
PPI	Plan Position Indicator
PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PSLR	Peak-to-Side Lobe Ratio
RCS	Radar Cross Section
RF	Radio Frequency
RFI	Radio Frequency Interference
RFM	Range Filter Map
RGPO	Range Gate Pull Off
RWR	Radar Warning Receiver
Rms	Root Mean Square
RWS	Radar Work Station
RX	Receiver
SAR	Synthetic Aperture Radar
SAW	Surface Acoustic Wave
SINR	Signal-to-Interference plus Noise Ratio
SLB	SideLobe Blanking
SLC	SideLobe Canceler
SNR	Signal-to-Noise Ratio
SOJ	Stand Off Jammer
SP	Self-Protection
SSJ	Self-Screening Jammer
STAP	Space-Time Adaptive Processing
STT	Single Target Tracker
ToA	Time of Arrival
TSI	Terrain Scattered Interference
TV-SAP	Time-Varying Spatial Adaptive Processing
TWS	Track While Scan
UHF	Ultra High Frequency
ULA	Uniform Linear Array
VGPO	Velocity Gate Pull Off
VHF	Very High Frequency
VLSI	Very Large Scale Integration

ACKNOWLEDGMENTS

The Author wishes to warmly thank his colleagues for their cooperation in this work: Dr. L. Timmoneri, Dr. L. Ortenzi, and Dr. E. Andreta (SELEX Sistemi Integrati, Italy), Dr. G. A. Fabrizio (DSTO, Australia), Dr. U. Nickel (FGAN, Germany), Prof. L. Chisci, Dr. A. Benavoli, and Dr. S. Romagnoli (University of Florence, Italy), Dr. M. Grazzini (Elettronica SpA, Italy), and Dr. S. Kogon (MIT, Lincoln Laboratory, U.S.).

REFERENCES

1. S. L. Johnston, "World War II ECCM history," suppl. to *IEEE Int. Radar Conf. Rec*., May 6–9, 1985, pp. 5.2–5.7.
2. A. E. Hoffmann-Heiden, "Anti-jamming techniques at the German AAA radars in World War II," suppl. to *IEEE Int. Radar Conf. Rec*., pp. 5.22–5.29, May 6–9, 1985.
3. D. C. Schleher, *Introduction to Electronic Warfare*, Norwood, MA: Artech House, Inc., 1986.
4. D. C. Schleher, *Electronic Warfare in the Information Age*, Norwood, MA: Artech House, Inc., 1999.
5. B. J. Slocumb and P. D. West, "ECM modeling for multitarget tracking and data association," in *Multitarget-Multisensor Tracking: Applications and Advances*, vol. III, Y. Bar-Shalom and W. D. Blair (eds.), Norwood, MA: Artech House, Inc., 2000, pp. 395–458.
6. F. Neri, *Introduction to Electronic Defense*, 2nd Ed., Norwood, MA: Artech House, Inc., 2001.
7. L. Nengjing and Z. Yi-Ting, "A survey of radar ECM-ECCM," *IEEE Trans*., vol. AES–31, no. 3, pp. 1110–1120, July 1995.
8. S. L. Johnston (ed.), *Radar Electronic Counter-Countermeasures*, Norwood, MA: Artech House, Inc., 1979.
9. Special Issue on electronic warfare, *IEE Proc*., vol. 129, pt. F, no. 3, pp. 113–232, June 1982.
10. W. A. Davis, "Principles of electronic warfare: Radar and EW," *Microwave J*., vol. 33, pp. 52–54, 56–59, February 1980.
11. L. B. Van Brunt, *The Glossary of Electronic Warfare*, Dunn Loring, VA: EW Engineering, Inc., 1984.
12. Department of Defense, Joint Chiefs of Staff, *Dictionary of Military and Associated Terms*, JCS Pub-1, September 1974.
13. L. B. Van Brunt, *Applied ECM*, vol. 1, Dunn Loring, VA: EW Engineering, Inc., 1978.
14. R. G. Wiley, *Electronic Intelligence: The Analysis of Radar Signals*, Norwood, MA: Artech House, Inc., 1985.
15. R. G. Wiley, *Electronic Intelligence: The Interception of Radar Signals*, Norwood, MA: Artech House, Inc., 1986.
16. R. G. Wiley, *ELINT: The Interception and Analysis of Radar Signals*, Norwood, MA: Artech House, Inc., 2006.
17. R. A. Poisel, *Electronic Warfare Target Location Methods*, Norwood, MA: Artech House, Inc., 2005.
18. E. P. Pace, *Detecting and Classifying Low Probability of Intercept Radar*, Norwood, MA: Artech House, Inc., 2003.
19. D. C. Schleher, "LPI radar: Fact or fiction," *IEEE AES Magazine*, vol. 21, no. 5, pp. 3–6, May 2006.
20. S. L. Johnston, "Philosophy of ECCM utilization," *Electron. Warfare*, vol. 7, pp. 59–61, May–June, 1975.
21. M. V. Maksimov, et al., *Radar Anti-Jamming Techniques*, Norwood, MA: Artech House, Inc., 1979. (Translated from Russian, Zaschita at Radiopomekh, Soviet Radio, 1976.)
22. D. Clifford Bell, "Radar countermeasures and counter-countermeasures," *Mil. Technol*., pp. 96–111, May 1986.
23. J. A. Adam and M. A. Fischetti, "Star Wars. SDI: The grand experiment," *IEEE Spectrum*, vol. 23, no. 9, pp. 34–46, September 1985.
24. S. J. Roome, "Digital radio frequency memory," *Electronic & Communication Engineering Journal*, pp. 147–153, August 1990.
25. J. W. Goodman and M. Silvestri, "Some effects of Fourier Domain Phase Quantization," *IBM J. Res. Develop*., pp. 478–484, September 1970.
26. M. Greco, F. Gini, and A. Farina, "Combined effect of phase and RGPO delay quantization on jamming signal spectrum," *Proc. of IEEE Int. Conf. on Radar*, Radar 2005, Washington, DC (USA), May 10–12, 2005, pp. 37–42.
27. S. D. Berger, "Digital radio frequency memory linear gate stealer spectrum," *IEEE Trans*., vol. AES–29, no. 2, pp. 725–735, April 2003.

28. G. V. Morris et al., "Principles of electronic counter-countermeasures," short lecture notes, Georgia Institute of Technology, 1999.
29. L. B. Van Brunt, *Applied ECM*, vol. 2, Dunn Loring, VA: EW Engineering, Inc., 1982.
30. P. J. Gros, D. C. Sammons, and A. C. Cruce, "ECCM Advanced Radar Test Bed (E/ARTB) systems definition," *IEEE Nat. Aerosp. Electron. Conf. NAECON 1986*, May 19–23, 1986, pp. 251–257.
31. M. A. Johnson and D. C. Stoner, "ECCM from the radar designer's view point," *Microwave J*., vol. 21, pp. 59–63, March 1978.
32. H. E. Schrank, "Low sidelobes phased-array and reflectors antennas," in *Aspects of Modern Radar*, E. Brookner (ed.), Norwood, MA: Artech House, Inc., 1988.
33. W. T. Patton, "Low Sidelobe Antennas for Tactical Radars," *IEEE Int. Radar Conf. Rec*., April 28–30, 1980, pp. 243–254.
34. A. Farina, *Antenna Based Signal Processing Techniques for Radar Systems*, Norwood, MA: Artech House, Inc., 1992.
35. F. J. Harrys, "On the use of windows for harmonic analysis with the Discrete Fourier Transform," *Proc. IEEE*, vol. 66, pp. 51–83, January 1978.
36. E. Brookner, "Trends in radar systems and technology to the year 2000 and beyond," in *Aspects of Modern Radar*, E. Brookner (ed.), Artech House, Inc., Norwood, MA, 1988.
37. E. Brookner, "Phased-array around the world. Progress and future trends," *IEEE Int. Symp. on Phased-Array Systems and Technology 2003*, Boston (USA), October 14–17, 2003, pp. 1–8.
38. M. Cicolani, A. Farina, E. Giaccari, F. Madia, R. Ronconi, and S. Sabatini, "Some phased-array systems and technologies in AMS," *IEEE Int. Symp. on Phased-Array Systems and Technology*, Boston (USA), October 14–17, 2003, pp. 23–30.
39. W. Kuhn, W. Sieprath, L. Timmoneri, and A. Farina, "Phased-array radar systems in support of the Medium Extended Air Defense System (MEADS)," *IEEE Int. Symp. on Phased-Array Systems and Technology*, Boston (USA), October 14–17, 2003, pp. 94–100.
40. A. R. Moore, D. M. Salter, and W. K. Stafford, "MESAR (Multi-Function, Electronically Scanned, Adaptive Radar)," *Proc. of Int. Conf. Radar 97*, Edinburgh, October 14–16, 1997, Publication no. 449, London, UK: IEE, pp. 55–59.
41. D. Giuli, "Polarization diversity in radars," *Proc. IEEE*, vol. 74, pp. 245–269, February 1986.
42. L. Maisel, "Performance of sidelobe blanking systems," *IEEE Trans*., vol. AES–4, no. 1, pp. 174–180, March 1968.
43. P. O. Arancibia, "A sidelobe blanking system design and demonstration," *Microwave J*., vol. 21, pp. 69–73, March 1978; reprinted in Ref. 8, 1979.
44. D. H. Harvey and T. L. Wood, "Designs for sidelobe blanking systems," *IEEE Int. Radar Conf. Rec*., April 1980, pp. 41–416.
45. M. O'Sullivan, "A comparison of sidelobe blanking systems, *IEE Int. Conf. Radar–87*, Conf. Pub. 281, London, UK, October 19–21, 1987, pp. 345–349.
46. A. Farina and F. Gini, "Calculation of blanking probability for the sidelobe blanking (SLB) for two interference statistical models," *IEEE Signal Processing Letters*, vol. 5, no. 4, pp. 98–100, April 1998.
47. A. Farina and F. Gini, "Blanking probabilities for SLB system in correlated clutter plus thermal noise," *IEEE Trans*., vol. SP–48, no. 5, pp. 1481–1485, May 2000.
48. A. Farina and F. Gini, "Design of SLB systems in presence of correlated ground clutter," *IEE Proc*., vol. 147, pt. F, no. 4, pp. 199–207, 2000.
49. A. De Maio, A. Farina, and F. Gini, "Performance analysis of the sidelobe blanking system for two fluctuating jammer models," *IEEE Trans*., vol. AES–41, no. 3, pp. 1082–1090, July 2005.
50. D. A. Shnidman and S. S. Toumodge, "Sidelobe blanking with integration and target fluctuation," *IEEE Trans*., vol. AES–38, no. 3, pp. 1023–1037, July 2002.
51. P. W. Howells, "Intermediate Frequency Sidelobe Canceler," U.S. Patent 3,202,990, August 24, 1965.
52. S. P. Applebaum, P. W. Howells, and C. Kovarik, "Multiple Intermediate Frequency Side-Lobe Canceler," U.S. Patent 4,044,359, August 23, 1977.

53. R. A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*, New York: John Wiley & Sons, 1980.
54. J. Hudson, *Adaptive Array Principles*, London: Peter Peregrinus Ltd., 1981.
55. R. Nitzberg, *Adaptive Signal Processing for Radar*, Norwood, MA: Artech House, Inc., 1992.
56. H. D. Griffiths, "A four-element VHF adaptive array processor," *Proc. 2nd IEE Int. Conf. on Antennas and Propagation*, IEE Conf. Pub. no. 195, pt.1, York (UK), April 13–16, 1981, pp. 185–189.
57. A. Farina, "Digital equalisation in adaptive spatial filtering: a survey," *Signal Processing*, Elsevier, vol. 83, no. 1, pp. 11–29, January 2003.
58. B. D. Carlson, L. M. Goodman, J. Austin, M. W. Ganz, and L. O. Upton, "An ultralow-sidelobe adaptive array antenna," *The Lincoln Laboratory Journal*, vol. 3, no. 2, pp. 291–310, 1990.
59. W. F. Gabriel, "Adaptive digital processing investigation of DFT sub-banding vs. transversal filter canceler," Naval Research Laboratory, NRL Report 8981, July 28, 1986, Washington, DC (USA).
60. A. Farina and R. Sanzullo, "Performance limitations in adaptive spatial filtering," *Signal Processing*, Elsevier, vol. 81, no.10, pp. 2155–2170, October 2001.
61. K. Gerlach, "The effects of IF bandpass mismatch errors on adaptive cancellation," *IEEE Trans*., vol. AES–26, no. 3, pp. 455–468, May 1990.
62. A. Farina, G. Golino, L. Timmoneri, and G. Tonelli, "Digital equalisation in adaptive spatial filtering for radar systems: Application to live data acquired with a ground-based phased-array radar," *Radar 2004*, Toulouse, France, October 19–21, 2004.
63. R. Fante, R. Davis, and T. Guella, "Wideband cancellation of multiple mainbeam jammers," *IEEE Trans*., vol. AP–44, no. 10, pp. 1402–1413, October 1996.
64. F. E. Churchill, G. W. Ogar, and B. J. Thompson, "The correction of I and Q errors in a coherent processor," *IEEE Trans*., vol. AES–17, no. 1, pp. 131–137, January 1981.
65. K. Gerlach, "The effect of I, Q mismatching errors on adaptive cancellation," *IEEE Trans*., vol. AES–28, no. 7, pp. 729–740, July 1992.
66. K. Gerlach and M. J. Steiner, "An adaptive matched filter that compensates for I, Q mismatch errors," *IEEE Trans*., vol. SP–45, no. 12, pp.3104–3107, December 1997.
67. A. Farina and L. Ortenzi, "Effect of ADC and receiver saturation on adaptive spatial filtering of directional interference," *Signal Processing*, Elsevier, vol. 83, no. 5, pp. 1065–1078, 2003.
68. A. Farina, R. Sanzullo, and L. Timmoneri, "Performance limitations and remedies in adaptive spatial filtering with timing errors," *Signal Processing*, Elsevier, vol. 82, no. 2, pp. 195–204, February 2002.
69. D. R. Morgan and A. Aridgides, "Adaptive sidelobes cancellation of wide-band multipath interference," *IEEE Trans*., vol. AP–33, no. 8, pp. 908–917, August 1985.
70. R. L. Fante, "Cancellation of specular and diffuse jammer multipath using a hybrid adaptive array," *IEEE Trans*., vol. AES–27, no. 10, pp. 823–837, September 1991.
71. A. Farina and L. Timmoneri, "Cancellation of clutter and e.m. interference with STAP algorithms. Application to live data acquired with a ground-based phased-array radar demonstrator," *Proc. of 2004 IEEE Radar Conf*., Philadelphia (USA), April 26–29, 2004, pp. 486–491.
72. A. Farina, L. Timmoneri, and R. Tosini, "Cascading SLB and SLC devices," *Signal Processing*, Elsevier, vol. 45, no. 2, pp. 261–266, 1995.
73. A. Farina and L. Timmoneri, "Systolic schemes for Joint SLB, SLC and adaptive phased-array," *Proc. of Int. Conf. on Radar*, Radar 2000, Washington, DC, USA, May 7–12, 2000, pp. 602–607.
74. L. Timmoneri, I. K. Proudler, A. Farina, and J. G. McWhirter, "QRD-Based MVDR algorithm for adaptive multipulse antenna array signal processing," *IEE Proc*., vol. 141, pt. F, no. 2, pp. 93–102, April 1994.
75. P. Bollini, L. Chisci, A. Farina, M. Giannelli, L. Timmoneri, and G. Zappa, "QR versus IQR algorithms for adaptive signal processing: performance evaluation for radar applications," *IEE Proc*., vol. 143, pt. F, no. 5, pp. 328–340, October 1996.
76. A. Farina and L. Timmoneri, "Real time STAP techniques," *Electronics & Communications Engineering Journal, Special Issue on STAP*, vol. 11, no.1, pp. 13–22, February 1999.

77. P. Kapteijin, E. Deprettere, L. Timmoneri, and A. Farina, "Implementation of the recursive QR algorithm on a 2*2 CORDIC test-board: a case study for radar application," *Proc. of the 25th European Microwave Conf*., Bologna (Italy), September 4–7, 1995, pp. 500–505.
78. A. D'Acierno, M. Ceccarelli, A. Farina, A. Petrosino, and L. Timmoneri, "Mapping QR decomposition on parallel computers: a study case for radar applications," *IEICE Trans. on Communications*, vol. E77–B, no. 10, pp. 1264–1271, October 1994.
79. A. Farina and L. Timmoneri, "Parallel processing architectures for STAP," in *Applications of Space-Time Adaptive Processing*, R. Klemm (ed.), London, UK, IEE Radar, Sonar and Navigation Series 14, 2004, pp. 265–302.
80. A. Farina, A. Averbouch, D. Gibor, L. Lescarini, S. Levit, S. Stefanini, and L. Timmoneri, "Multi-channel radar: Advanced implementation technology and experimental results," *Proc. of Int. Radar Symp*., IRS2005, Berlin (Germany), September 6–8, 2005, pp. 317–329.
81. C. M. Rader, "Wafer scale integration of a large scale systolic array for adaptive nulling," *The Lincoln Laboratory Journal*, vol. 4, no. 1, pp. 3–29, 1991.
82. C. M. Rader, "VLSI systolic array for adaptive nulling," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 29–49, July 1996.
83. S. P. Applebaum, "Adaptive arrays," Syracuse University Research Corporation Rept. SPL TR 66–1, 1966. This report is reproduced in *IEEE Trans*., vol. AP–24, pp. 585–598, September 1976.
84. L. E. Brennan and I. S. Reed, "Theory of adaptive radar," *IEEE Trans*., vol. AES-9, no. 1, pp. 237–252, March 1973.
85. B. Wardrop, "The role of digital processing in radar beamforming," *GEC J. Res*., vol. 3, no. 1, pp. 34–45, 1985.
86. P. Valentino, "Digital beamforming: new technology for tomorrow's radars," *Def. Electron*., pp. 102–107, October 1984.
87. H. Steyskal, "Digital beamforming antennas: an introduction," *Microwave J*., pp. 107–124, January 1987.
88. B. Cantrell, J. de Graaf, L. Leibowitz, E. Willwerth, G. Meurer, C. Parris, and R. Stapleton, "Development of a Digital Array Radar (DAR)," *Proc. of IEEE Radar Conf. 2001*, Atlanta (Georgia), May 1–3, 2001 pp. 157–162.
89. M. Zatman, "Digitization requirements for digital radar arrays," *IEEE Radar Conf. 2001*, Atlanta (Georgia), May 1–3, 2001, pp. 163–168.
90. I. S. Reed, "A brief history of adaptive arrays," *Subdury/Wayland Lecture Series*, Raytheon Div. Education, notes 23, October 1985.
91. D. Etter, A. Steinhardt, and S. Stoner, "Least squares adaptive processing in military applications," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 66–73, May 2002. On occasion of the 2001 B. Franklin Medal awarded to B. Widrow for pioneering work on adaptive signal processing.
92. S. Haykin and A. Steinhardt, *Adaptive Radar Detection and Estimation*, New York: John Wiley & Sons, Inc., 1992.
93. S. T. Smith, "Adaptive Radar," in *Wiley Encyclopedia of Electrical and Electronic Engineering*, J. G. Webster (ed.), vol. 1, New York: Wiley, 1999 (updated 13 July 2007), pp. 263–289.
94. A. Farina, C. H. Gierull, F. Gini, and U. Nickel (eds.), Special Issue "New trends and findings in antenna array processing," *Signal Processing*, Elsevier, vol. 84, no. 9, pp. 1477–1688, September 2004.
95. J. Ward, "Space-time adaptive processing for airborne radar," MIT Lincoln Laboratory Technical Report TR–1015, December 13, 1994.
96. R. Klemm, *Principles of Space-Time Adaptive Processing*, 3rd Ed., London, UK: IET Radar, Sonar and Navigation Series 21, 2006.
97. R. Klemm (ed.), *Applications of Space-Time Adaptive Processing*, London, UK: IEE Radar, Sonar and Navigation, Series 14, 2004.
98. J. R. Guerci, *Space-Time Adaptive Processing for Radar*, Norwood, MA: Artech House, Inc., 2003.
99. B. Testa and V. Vannicola, "The physical significance of the eigenvalues in adaptive arrays," *Digital Signal Processing*, vol. 15, pp. 91–96, 1995.

100. B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans*., vol. AES–24, no. 3, pp. 397–401, July 1988.

101. A. Farina, P. Langsford, G. C. Sarno, L. Timmoneri, and R. Tosini, "ECCM techniques for a rotating, multifunction, phased-array radar," *Proc. of the 25th European Microwave Conf*, Bologna (Italy), September 4–7, 1995, pp. 490–495.

102. W. D. Wirth, *Radar Techniques Using Array Antennas*, London, UK: IEE Radar, Sonar, Navigation and Avionics, Series 10, 2001.

103. J. B. Hoffman and B. L. Gabelach, "Four-channel monopulse for main beam nulling and tracking," *Proc. of IEEE National Radar Conf. NATRAD '97*, Syracuse, New York, May 13–15, 1997, pp. 94–98.

104. A. Farina, P. Lombardo, and L. Ortenzi, "A unified approach to adaptive radar processing with general antenna array configuration," Special Issue on "New trends and findings in antenna array processing for radar," *Signal Processing*, Elsevier, vol. 84, no. 9, pp. 1593–1623, September 2004.

105. R. C. Davis, L. E. Brennan, and I. S. Reed, "Angle estimation with adaptive arrays in external noise field," *IEEE Trans*., vol. AES–12, no. 2 pp. 179–186, March 1976.

106. P. Langsford A. Farina, L. Timmoneri, and R. Tosini, "Monopulse direction finding in presence of adaptive nulling," presented at IEE Colloquium on Advances in Adaptive Beamforming, Romsey, UK, June 13, 1995.

107. F. C. Lin and F. F. Kretschmer, "Angle measurement in the presence of mainbeam interference," *Proc. of IEEE 1990 Int. Radar Conf*., Arlington (VA), USA, May 7–10, 1990, pp. 444–450.

108. U. Nickel, "Monopulse estimation with adaptive arrays." *IEE Proc*., vol. 130, pt. F, no. 5, pp. 303–308, October 1993.

109. M. Valeri, S. Barbarossa, A. Farina, and L. Timmoneri, "Monopulse estimation of target DoA in external fields with adaptive arrays," *IEEE Symp. of Phased-Array Systems and Technology*, Boston (MA), USA, October 15–18, 1996, pp. 386–390.

110. U. Nickel, "Performance of corrected adaptive monopulse estimation," *IEE Proc*., vol. 146, pt. F, no. 1, pp. 17–24, February 1999.

111. J. Worms, "Monopulse estimation and SLC configurations," *Proc. of IEEE Radar Conf*. 1998, Dallas, TX, May 11–14, 1998, pp. 56–61.

112. U. Nickel, "Overview of generalized monopulse estimation," *IEEE AES Magazine*, vol. 21, no. 6, part 2 of 2, pp. 27–56, June 2006.

113. A. Farina, G. Golino, and L. Timmoneri, "Maximum likelihood estimator approach for the estimation of target angular coordinates in presence of main beam interference: Application to live data acquired with a ground-based phased-array radar," *Proc. of IEEE 2005 Int. Radar Conf*., Alexandria (VA), USA, May 9–12, 2005, pp. 61–66.

114. A. Farina, G. Golino, and L. Timmoneri, "Maximum likelihood estimate of target angular coordinates under main beam interference: Application to recorded live data," in *Advances in Direction-of-Arrival Estimation*, S. Chandran (ed.), Norwood, MA: Artech House, Inc., 2006, pp. 285–303.

115. J. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans*., vol. AES–28, no. 1, pp. 208–216, January 1982.

116. A. Farina, G. Golino, and L. Timmoneri, "Comparison between LS and TLS in adaptive processing for radar systems," *Proc. of IEE*, vol. 150, pt. F, no. 1, pp. 2–6, February 2003.

117. A. Farina and L. Timmoneri, "Cancellation of clutter and e.m. interference with STAP algorithm. Application to live data acquired with a ground-based phased array radar," *Proc. of IEEE 2004 Radar Conf*., Philadelphia (USA), April 26–29, 2004, pp. 486–491.

118. A. Farina, G. Golino, S. Immediata, L. Ortenzi, and L. Timmoneri, "Techniques to design subarrays for radar phased-array antennas," *IEE Int. Conf. on Antennas and Propagation (ICAP) 2003*, March 31– April 3, 2003, pp. 17–23.

119. U. Nickel, "Sub-array configurations for digital beamforming with low sidelobes and adaptive interferences suppression," *Proc. IEEE 1995 Int. Radar Conf*., Alexandria (VA), USA, May 8–11, 1995, pp. 714–719.

120. U. Nickel, "Monopulse estimation with sub-array output adaptive beam forming and low side lobe sum and difference beams," *IEEE Symp. on Phased-Array Systems and Technology*, Boston (MA), USA, October 15–18, 1996, pp. 283–288.

121. E. Brookner and J. M. Howells, "Adaptive-Adaptive Array Processing," *IEE Int. Conf. Radar–87*, Conf. Pub. 281, London, October 19–21, 1987, pp. 257–263.

122. L. W. Dicken, "The use of null steering in suppressing main beam interference," *IEE Int. Conf. Radar–77*, Conf. Pub. 155, London, October 25–28, 1977, pp. 226–231.

123. W. F. Gabriel, "Spectral analysis and adaptive array superresolution techniques," *Proc. IEEE*, vol. 68, pp. 654–666, June 1980.

124. U. Nickel, "Fast subspace methods for radar applications," in *Advanced Signal Processing: Algorithms, Architectures and Implementation VII*, F. T. Luk (ed.), SPIE Proc. Series vol. 3162 (Conf. Rec. SPIE San Diego 1997), pp. 438–448.

125. U. Nickel, "Aspects of implementing superresolution methods into phased array radar," *Int. Journal Electronics and Communications* (AEÜ), vol. 53, no. 6, pp. 315–323, 1999.

126. U. Nickel, "Spotlight MUSIC: Superresolution with sub-arrays with low calibration effort," *IEE Proc*., vol. 149, pt. F, no. 4, pp. 166–173, August 2002.

127. U. Nickel, "Superresolution and jammer suppression with broadband arrays for multi-function radar," Chapter 16 in *Applications of Space-Time Adaptive Processing*, R. Klemm (ed.), London: IEE, 2004, pp. 543–599.

128. H. Lee, "Eigenvalues and eigenvectors of covariance matrices for signal closely spaced in frequency," *IEEE Trans*., vol. SP–40, no. 10, pp. 2518–2535, October 1992.

129. Special Issue on Superresolution, *The Lincoln Laboratory Journal*, vol. 10, no. 2, pp. 83–222.

130. S. T. Smith, "Statistical resolution limits and complexified Cramer-Rao bound," *IEEE Trans*., vol. SP–53, no. 5, pp. 1597–1609, May 2005.

131. D. K. Barton, *Radar*, vol. 6, *Frequency Agility and Diversity*, Norwood, MA: Artech House, Inc., 1977.

132. B. Bergkvist, "Jamming frequency agile radars," *Def. Electron*., vol. 12, pp. 75.78–81.83, January 1980.

133. S. Strappaveccia, "Spatial jammer suppression by means of an automatic frequency selection device," *IEE Int. Conf. Radar–87*, Conf. Pub. 281, London, October19–21, 1987, pp. 582–587.

134. C. H. Gager, "The impact of waveform bandwidth upon tactical radar design," *IEE Int. Conf. Radar-82*, London, October 18–20, 1982, pp. 278–282.

135. M. I. Skolnik, G. Linde, and K. Meads, "Senrad: An advanced wideband air surveillance radar," *IEEE Trans*., vol. AES–37, no. 4, pp. 1163–1175, October 2001.

136. B. L. Lewis, F. F. Kretschmer, and W. W. Shelton, *Aspects of Radar Signal Processing*, Norwood, MA: Artech House, Inc., 1986.

137. N. Levanon and E. Mozeson, *Radar Signals*, New York: John Wiley & Sons, Inc., 2004.

138. G. Petrocchi, S. Rampazzo, and G. Rodriguez, "Anticlutter and ECCM design criteria for a low coverage radar," *Proc. Int. Conf. Radar*, Paris, France, December 4–8, 1978, pp. 194–200.

139. V. G. Hansen and A. J. Zottl, "The detection performance of the Siebert and Dicke-Fix CFAR detectors," *IEEE Trans*., vol. AES–7, pp. 706–709, July 1971.

140. S. L. Johnston, "Radar electronic counter-countermeasures against chaff," *Proc. Int. Conf. Radar*, Paris, France, May 1984, pp. 517–522.

141. M. I. Skolnik, *Introduction to Radar Systems*, 3rd Ed., New York: McGraw-Hill, 2001.

142. A. Farina and F. A. Studer, "A review of CFAR detection techniques in radar systems," *Microwave Journal*, pp. 115–128, September 1986.

143. E. Conte and A. De Maio, "Mitigation techniques for non-gaussian sea clutter," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 284–302, April 2004.

144. E. Conte, A. De Maio, A Farina, and G. Foglia, "CFAR behavior of adaptive detectors: an experimental analysis," *IEEE Trans*., vol. AES–41, no. 1, pp. 233–251, January 2005.

145. M. C. Wicks, W. J. Baldygo, and R. D. Brown, "Expert System Application to Constant False Alarm Rate (CFAR) Processor," U.S. Patent 5, 499, 030, March 12, 1996.

146. A. Farina (ed.), *Optimised Radar Processors*, London: Peter Peregrinus, Ltd., 1987.
147. E. Fong, J. A. Walker, and W. G. Bath, "Moving target indication in the presence of radio frequency interference," *Proc. IEEE 1985 Int. Radar Conf*., Arlington (VA), USA, May 6–9, 1985, pp. 292–296.
148. L. B.Van Brunt, "Pulse-compression radar: ECM and ECCM," *Def. Electron*., vol. 16, pp. 170–185, October 1984.
149. H. Kushel, "VHF/UHF. Part 1: characteristics," *Electronics & Communications Engineering Journal*, vol. 14, no. 2, pp. 61–72, April 2002.
150. R. J. Galejs, "Volume surveillance radar frequency selection," *Proc. of IEEE 2000 Int. Radar Conf*., Alexandria (VA), USA, May 7–12, 2000, pp. 187–192.
151. H. Kushel, "VHF/UHF. Part 2: operational aspects and applications," *Electronics & Communications Engineering Journal*, vol. 14, no. 3, pp. 101–111, June 2002.
152. W. N. Dawber and N. M. Harwood, "Comparison of doppler clutter cancellation techniques for naval multi-function radars," *IEE Int. Conf. Radar 2002*, Conf. Pub. No. 490, Edinburgh, UK, 15-17 October 2002, pp. 424-428.
153. A. I. Leonov and K. J. Fomichev, *Monopulse Radar*, Norwood, MA: Artech House, Inc., 1987.
154. S. L. Johnston, "Tracking radar electronic counter-countermeasures against inverse gain jammers," *IEE Int. Conf. Radar–82*, Conf. Pub. 216, London, October 1982, pp. 444–447.
155. B. L. Lewis and D. H. Howard, "Security Device," U.S. Patent, 4, 006, 478, February 1, 1977, filed August 15, 1958.
156. R. L. Fante and J. A. Torres, "Cancellation of diffuse jammer multipath by an airborne adaptive radar," *IEEE Trans*., vol. AES–31, no. 2, pp. 805–820, April 1995.
157. S. Kogon, "Algorithms for mitigating terrain-scattered interference," *Electronics & Communications Engineering Journal*, vol. 11, no. 1, pp. 49–56, February 1999.
158. S. Bjorklund and A. Nelander, "Theoretical aspects on a method for terrain scattered interference mitigation in radar," *Proc. of IEEE 2000 Int. Radar Conf*., Alexandria (VA), USA, May 9–12, 2005, pp. 663–668.
159. Y. Abramovich, S. J. Anderson, and A. Y. Gorokov, "Stochastically constrained spatial and spatio-temporal adaptive processing for non-stationary hot clutter cancellation," Chapter 17 in *Applications of Space-Time Adaptive Processing*, R. Klemm (ed.), London: IEE Radar, Sonar and Navigation, Series 14, 2004, pp. 603–697.
160. W. D. Blair, G. A. Watson, T. Kirubarajan, and Y. Bar-Shalom, "Benchmark for radar allocation and tracking in ECM," *IEEE Trans*., vol. AES–34, no.4, pp.1097–1114, 1998.
161. T. Kirubarajan, Y. Bar-Shalom, W. D. Blair, and G. A. Watson, "IMMPDAF for radar management and tracking benchmark with ECM," *IEEE Trans*., vol. AES–34, no.4, pp.1115–1134, 1998.
162. H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans*., vol. AC–33, no. 8, pp. 780–783, August 1988.
163. A. Moreira, "Improved multilook techniques applied to SAR and SCANSAR imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 29, no. 4, pp. 529–534, July 1991.
164. W. Goj, *Synthetic Aperture Radar and Electronic Warfare*, Dedham, MA: Artech House, Inc., 1989.
165. C. J. Condley, "The potential vulnerability to increased background noise of synthetic aperture radar in the maritime environment," *IEE Colloquium on Synthetic Aperture Radar*, November 29, 1989, pp. 10/1–10/5.
166. C. J. Condley, "Some system considerations for electronic countermeasures to synthetic aperture radar," *IEE Colloquium on Electronic Warfare Systems*, January 14, 1991, pp. 8/1–8/7.
167. M. Soumekh, "SAR-ECCM using phased-perturbed LFM chirp signals and DRFM repeat jammer penalizer," *IEEE Trans*., vol. AES–42, no. 1, pp. 191–205, January 2006.
168. P. Hyberg, "Assessment of modern coherent jamming methods against synthetic aperture radar (SAR)," *Proc. of EUSAR '98, European Conf. on Synthetic Aperture Radar*, Friedrichshafen, Germany, May 25–27, 1998, pp. 391–394.
169. C. Boesswetter, "ECCM effectiveness of a low sidelobe antenna for SAR ground mapping," AGARD AVP Symp. "*Multifunction Radar for Airborne Applications*," Toulouse, 1985.

170. A. Farina and P. Lombardo, "SAR ECCM using adaptive antennas," *Proc. of IEEE Long Island Section, Adaptive Antenna Systems Symp*., Long Island, USA, November 1994, pp. 79–84.
171. J. H. Ender, "Anti-jamming adaptive filtering for SAR imaging," *Proc. of IRS '98, Int. Radar Symp*., Munich, Germany, September 15–17, 1998, pp. 1403–1413.
172. J. A. Torres, R. M. Davis, J. D. R. Kramer, and R. L. Fante, "Efficient wideband jammer nulling when using stretch processing," *IEEE Trans*., vol. AES–36, no. 4, pp. 1167–1178, October 2000.
173. L. Rosenberg and D. Gray, "Anti-jamming techniques for multi-channel SAR imaging," *IEE Proc*., pt. F, vol. 133, no. 3, pp. 234–242, June 2006.
174. P. E. Pace, D. J. Fouts, S. Ekestrom, and C. Karow, "Digital false target image synthesizer for countering ISAR," *IEE Proc*., pt. F, vol. 149, no. 5, pp. 248–257, October 2002.
175. P. E. Pace, D. J. Fouts, and D. P. Zulaica, "Digital image synthesizer: Are enemy sensors really seeing what's there?," *IEEE Aerospace and Electronic Systems Magazine*, vol. 24, no. 2, pp. 3–7, February 2006.
176. L. Sevgi, A. Ponsford, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars, part 1: Theoretical background and numerical simulations," *IEEE Antennas and Propagation Magazine*, vol. 43, no. 5, pp. 28–43, October 2001.
177. A. Ponsford, L. Sevgi, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars, part 2: Operational status and system performance," *IEEE Antennas and Propagation Magazine*, vol. 43, no. 5, pp. 52–63, October 2001.
178. G. A. Fabrizio, "Space-time characterization and adaptive processing of ionospherically-propagated HF signals," Ph.D. dissertation, Adelaide University, Australia, July 2000.
179. G. A. Fabrizio, D. A. Gray, and M. D. Turley, "Experimental evaluation of adaptive beamforming methods and interference models for high frequency over-the-horizon radar," *Multidimensional Systems and Signal Processing – Special Issue on Radar Signal Processing Techniques*, vol.14, no. 1/2/3, pp. 241–263, January–July 2003.
180. G. A. Fabrizio, Y. I. Abramovich, S. J. Anderson, D. A. Gray, and M. D. Turley, "Adaptive cancellation of nonstationary interference in HF antenna arrays," *IEE Proc*., vol. 145, pt. F, no. 1, pp. 19–24, February 1998.
181. Y. I. Abramovich, A. Y. Gorokhov, V. N. Mikhaylyukov, and I. P. Malyavin, "Exterior noise adaptive rejection for OTH radar implementations," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 1994, ICASSP'94*, Adelaide (Australia), 1994, pp. 105–107.
182. S. J. Anderson, Y. I. Abramovich, and G. A. Fabrizio, "Stochastic constraints in non stationary hot clutter cancellation," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 1997, ICASSP–97*, Munich, Germany, vol. 5, pp. 21–24, April 1997, vol. 5, pp. 3753 – 3756.
183. Y. I. Abramovich, N. Spencer, and S. J. Anderson, "Stochastic constraints method in non stationary hot clutter cancellation–part 1: Fundamentals and supervised training applications," *IEEE Trans*., AES–34, no. 4, pp. 1271–1292, 1998.
184. Y. I. Abramovich, N. Spencer, and S. J. Anderson, "Stochastic constraints method in non stationary hot clutter cancellation–part 2: Unsupervised training applications," *IEEE Trans*., vol. AES–36, no. 1, pp. 132–150, 2000.
185. G. A. Fabrizio, A. B. Gershman, and M. D. Turley, "Robust adaptive beamforming for HF surface wave over-the-horizon," *IEEE Trans*., vol. AES–40, no. 2, pp. 510–525, April 2004.
186. G. A. Fabrizio, A. Farina, and M. D. Turley, "Spatial adaptive subspace detection in OTH radar," *IEEE Trans*., vol. AES–39, no. 4, pp. 1407–1428, October 2003.
187. A. Farina, G. A. Fabrizio, W. L. Melvin, and L. Timmoneri, "Multichannel array processing in radar: State of the art, hot topics and way ahead," *Proc. Sensor Array and Multichannel Signal Processing IEEE Workshop* (invited paper), Sitges, Spain, July 18–21, 2004, pp. 11–19.
188. G. A. Fabrizio, G. J. Frazer, and M. D. Turley, "STAP for Clutter and Interference Cancellation in a HF Radar System," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing 2006*, ICASSP 2006, Toulouse, France, May 2006.
189. D. K. Barton, *Radar System Analysis and Modeling*, Norwood, MA: Artech House, Inc., 2005.

190. S. L. Johnston, "The ECCM improvement factor (EIF): illustration examples, applications, and considerations in its utilization in radar ECCM performance assessment," *Int. Conf. Radar*, Nanjing (China), November 4–7, 1986, pp. 149–154.

191. J. Clarke and A. R. Subramanian, "A game theory approach to radar ECCM evaluation," *Proc. of IEEE 1985 Int. Radar Conf.*, Arlington (VA), USA, May 6–9, 1985, pp. 197–203.

192. L. Nengjing, "Formulas for measuring radar ECCM capability," *IEE Proc*., vol. 131, pt. F, pp. 417–423, July 1984.

193. L. Nengjing, "ECCM efficacy assessment in surveillance radar analysis and simulation," *IRS '98, Int. Radar Symp*., Munich, Germany, September 15–17, 1998, pp. 1415–1419.

194. D. H. Cook, "ECM/ECCM systems simulation program, electronic and aerospace systems record," *IEEE Conv. Rec. EASCON '68*, September 9–11, 1968, pp. 181–186.

195. S. Watts, H. D. Griffiths, J. R. Hollaway, A. M. Kinghorn, D. G. Money, D. J. Price, A. M. Whitehead, A. R. Moore, M. A. Wood, and D. J. Bannister, "The specification and measurement of radar performance," *IEE Int. Conf. Radar 2002*, Conf. Pub. no. 490, Edinburgh, UK, October 15–17, 2002, pp. 542–546.

196. F. A. Studer, M. Toma, and F. Vinelli, "Modern software tools for radar performance assessment," *Proc. of IRS '98, Int. Radar Symp*., Munich, Germany, September 15–17, 1998, pp. 1079–1090.

197. A. G. Huizing and A. Theil, *CARPET 2.11 Software + User Manual*, The Hague, The Netherlands: TNO Defense, Security and Safety, 2004.

Chapter 25
Radar Digital Signal Processing

James J. Alter
Jeffrey O. Coleman

Naval Research Laboratory

25.1 INTRODUCTION

The exponential growth in digital technology since the 1980s, along with the corresponding decrease in its cost, has had a profound impact on the way radar systems are designed. More and more functions that historically were implemented in analog hardware are now being performed digitally, resulting in increased performance and flexibility and reduced size and cost. Advances in analog-to-digital converter (ADC) and digital-to-analog converter (DAC) technologies are pushing the border between analog and digital processing closer and closer to the antenna.

For example, Figure 25.1 shows a simplified block diagram of the receiver front end of a typical radar system that would have been designed around 1990. Note that this system incorporated analog pulse compression (PC). It also included several stages of analog downconversion, in order to generate baseband in-phase (I) and quadrature (Q) signals with a small enough bandwidth that the ADCs of the day could sample them. The digitized signals were then fed into digital doppler/MTI and detection processors.

By contrast, Figure 25.2 depicts a typical digital receiver for a radar front end. The RF input usually passes through one or two stages of analog downconversion to generate an Intermediate Frequency (IF) signal that is sampled directly by the ADC. A digital downconverter (DDC) converts the digitized signal samples to complex form at a lower rate for passing through a digital pulse compressor to backend processing. Note that the output of the ADC has a slash through the digital signal line with a letter above. The letter depicts the number of bits in the digitized input signal and represents the maximum possible dynamic range of the ADC. As will be described later, the use of digital signal processing (DSP) can often improve the dynamic range, stability, and overall performance of the system, while reducing size and cost, compared to the analog approach.

This chapter will provide a high-level outline of some of the major digital processing techniques for radar systems that have become practical since the Second Edition of this *Handbook* was published, as well as some design tradeoffs that need to be considered.

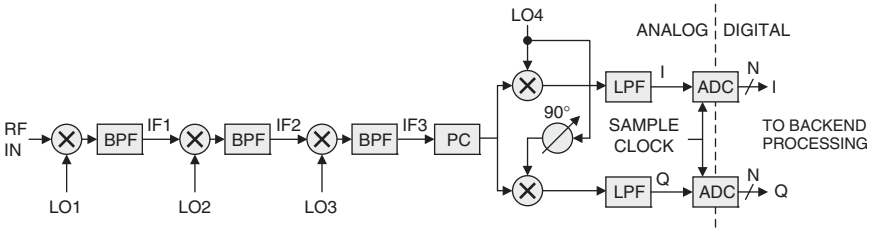


FIGURE 25.1 Typical radar receiver front-end design from 1990

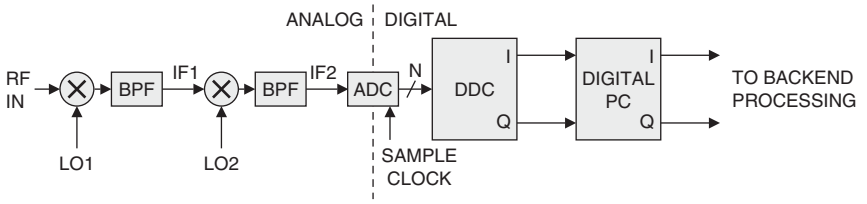


FIGURE 25.2 Typical digital receiver front end

25.2 RECEIVE CHANNEL PROCESSING

Major advances in analog-to-digital converter and digital component technology have transformed the receiver front ends of radar systems, providing higher performance at lower cost. This section will describe how these new technologies are being applied to radar systems and the benefits they bring to system performance.

Signal Sampling Basics. Digital signal processors are sampled signal systems. *Sampling* is the process by which a continuous (analog) signal is measured at regular intervals of time (the *sampling interval*), producing a sequence of discrete numbers (samples) that represents the values of the signal at the sampling instants. The *sampling frequency* is the inverse of the sampling interval and is typically designated f_s. Sampled systems are subject to the Nyquist limit,[1] which lower bounds the sampling rate at which reconstruction of the unsampled signal from its samples is possible without corruption by *aliasing*, the overlapping of spectral components. The bound, termed the *Nyquist frequency* or *Nyquist rate,* is equal to the two-sided signal bandwidth B, the bandwidth considering components at both positive and negative frequencies. Sampling below the Nyquist rate always results in aliasing, but sampling above it does not guarantee alias-free operation. We will see that for bandpass signals a sampling rate higher than Nyquist may be required to avoid aliasing in some situations.

The Nyquist rate is often said to be twice the signal bandwidth, but that refers to a one-sided bandwidth, positive frequencies only, of a real signal. Our definition refers to the two-sided bandwidth, both positive and negative frequencies, of a signal that, in general, is complex with a real signal as a special case.

Is the two-sided bandwidth always twice the one-sided bandwidth? For complex signals in general, no, but for real signals in particular, yes. Here's why: any signal, real or complex, when expressed as a Fourier integral (inverse Fourier transform) is

seen to be a combination of spectral components of the form $A\,e^{j2\pi ft}$. Sampled signals have $t = nT$ with T a sampling interval and n an integer time, but sampled or not, the basic component form is the same. And either way, complex amplitude A is a function of frequency f, but let's write A instead of $A(f)$ for simplicity.

In these terms then, what's special about real signals is that an easily derived Fourier-transform property requires their Fourier components to occur in conjugate pairs, so that if there is a component $A\,e^{j2\pi ft}$ at frequency f with complex amplitude A, there is also a component $A^*\,e^{-j2\pi ft}$ at frequency $-f$ with the complex conjugate A^* of that complex amplitude. If a band of positive frequencies from f_1 to f_2 is occupied by spectral components, the corresponding band of negative frequencies from $-f_2$ to $-f_1$ will be occupied by spectral components also, so the two-sided bandwidth must be twice the one-sided bandwidth.

Real signals have spectral components in conjugate pairs because by using complex amplitude expressed in polar form as $A = re^{j\theta}$,

$$A\,e^{j2\pi ft} + A^*\,e^{-j2\pi ft} = 2\,\mathrm{Re}\{Ae^{j2\pi ft}\} = 2\,\mathrm{Re}\{r\,e^{j\theta}e^{j2\pi ft}\}$$
$$= 2r\,\mathrm{Re}\{e^{j(2\pi ft+\theta)}\} = 2r\cos(2\pi ft+\theta)$$

The imaginary parts of the conjugate spectral components have canceled to reveal that those components together indeed represent a real signal, a sinusoid with amplitude and phase specified by the magnitude and angle of the complex amplitude. The latter relationship is so much a part of the engineering culture that the terms *amplitude* and *phase* are commonly, if imprecisely, used to refer to the magnitude and angle of a complex signal at an instant in time.

The following figures illustrate the origin of the Nyquist rate. Imagine that a real signal with a *lowpass* signal spectrum of two-sided bandwidth B is plotted on a long piece of paper, as shown in Figure 25.3*a*. In the figure, the positive-frequency spectral components of the signal are darkly shaded, and the negative-frequency components are lightly shaded. To see the effect of sampling this signal at Nyquist rate B, the long sheet is cut into smaller sheets, with the first cut at zero frequency and subsequent cuts at sample-rate (B, in this case) intervals in positive and negative frequency. The sheets are stacked one on top of the other as shown on the left side of Figure 25.3*b*, and the resulting portion of the sampled signal spectrum from 0 to the sampling rate of B is generated by adding the spectra of the stacked pages together, as shown on the right.

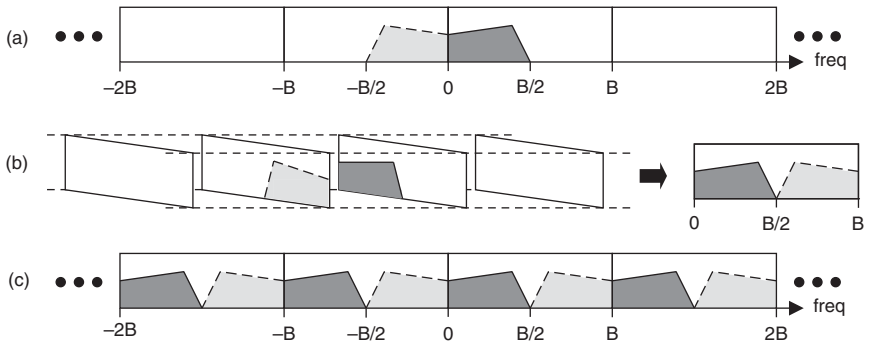


FIGURE 25.3 (*a*) Bandlimited, real signal spectrum before sampling, (*b*) portion of sampled spectrum from 0 to B, and (*c*) full sampled signal spectrum

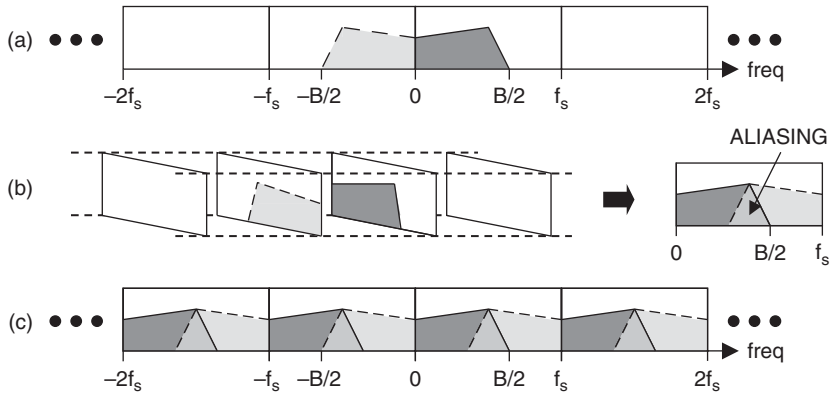


FIGURE 25.4 (*a*) Bandlimited lowpass signal spectrum before sampling, (*b*) aliased lowpass signal spectrum after sampling at rate $f_s < B$, and (*c*) aliased sampled signal spectrum

Note that the lightly shaded negative-frequency portion of the spectrum now appears on the right of the sampled spectrum and doesn't overlap the darker positive-frequency portion. As long as the two portions of the sampled signal don't overlap, the signal is not aliased. The full sampled-signal spectrum is obtained by laying copies of this page end-to-end, as shown in Figure 25.3*c*, producing copies of the 0 to *B* portion of the sampled signal spectrum at *B* intervals.

Figure 25.4 shows the result of sampling below the Nyquist rate. Figure 25.4*a* shows the same bandlimited signal as the previous example, but this time it is sampled at some rate that is less than Nyquist rate *B*. The resulting sampled spectrum, shown in Figure 25.4*b* and *c*, contains overlapped, or aliased, spectral components that add and represent corruption of the signal.

Figure 25.5 repeats this Nyquist analysis for a *bandpass signal*—a signal not containing spectral components at or near 0 Hz. Figure 25.5*a* shows a real bandpass signal with two-sided bandwidth *B* and composed of positive-frequency and negative-frequency spectral components, each of bandwidth *B*/2, that are complex-conjugated mirror images. The Nyquist rate is the signal's two-sided bandwidth irrespective of the particular portion of the spectrum in which the signal resides. Therefore, for this signal the Nyquist frequency is *B* even though the signal contains components at actual frequencies greater than *B*. Figure 25.5*b* shows the result of sampling this signal at

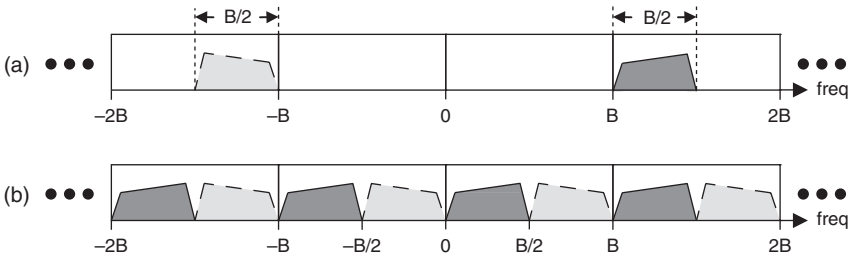


FIGURE 25.5 (*a*) Bandlimited, real passband signal spectrum before sampling and (*b*) signal spectrum after sampling

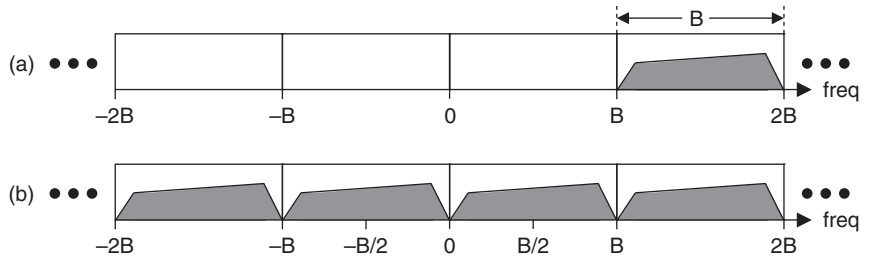


FIGURE 25.6 (*a*) Non-real signal spectrum before sampling by Nyquist frequency, *B*, and (*b*) signal spectrum after sampling

the Nyquist bound. The sampled spectra of the two portions of the signal do not overlap; the sampled signal is not aliased. As will be described in more detail later in the chapter, this technique, *bandpass sampling,* is a powerful tool that allows a relatively high-frequency signal to be sampled by a relatively low-performance digitizer, which can result in considerable cost savings.

Figure 25.6*a* shows the spectrum of a more general complex signal of bandwidth *B* before sampling. Note that this signal does not possess complex-conjugate spectral symmetry. The signal spectrum after sampling by its Nyquist frequency *B* is shown in Figure 25.6*b*. There is no aliasing.

The Nyquist rate is a *minimum* sampling frequency for a signal, a bound, and meeting the bound is necessary, but not sufficient, to ensure that no aliasing occurs. Consider the case presented in Figure 25.7*a*, which is the same bandlimited bandpass signal shown in Figure 25.5, but shifted in frequency so that it doesn't begin exactly at *B.* The sampled signal spectrum in Figure 25.7*b* shows that, although the sampling rate satisfies the Nyquist bound, the sampled signal is still aliased. To solve this problem the signal could be moved to a different center frequency before sampling or the sampling rate could be increased. The system designer must always develop the frequency plan of a sampling system carefully to determine an appropriate sampling frequency and to ensure that aliasing does not occur. A full treatment of this subject is presented by Lyons.[2]

In an actual system, before sampling the signal is typically passed through an *anti-aliasing filter*, which is an analog lowpass or bandpass filter that places an upper limit on the signal bandwidth. The filter needs to provide enough stopband rejection that any aliased components are insignificant. Of course, practical filters do not have

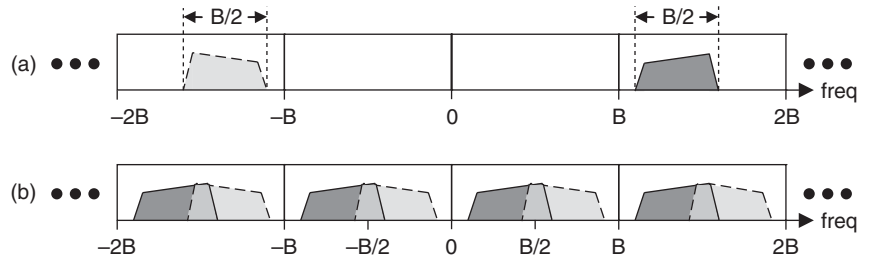


FIGURE 25.7 (*a*) Bandlimited, real passband signal spectrum before sampling and (*b*) signal spectrum after sampling

passbands extending right up to their stopband edges, so the widths of intervening transition bands must be counted as part of two-sided signal bandwidth B for purposes of determining the Nyquist rate, as the filter output may contain components in these transitions bands that could otherwise result in significant aliasing.

Digital Downconversion (DDC). The application of digital technology to IQ demodulation, which is just downconversion of an IF signal to a complex baseband, has greatly improved the performance of coherent systems. Here, we explore two forms of such digital downconversion, a general form that is structurally parallel to traditional analog downconversion and a restricted form, direct digital downconversion, which is more economical when it is applicable.

Analog Downconversion and Sampling. The general approach to digital downconversion derives from analog downconversion and sampling, as illustrated in the frequency domain in Figure 25.8. The spectra on the first and = lines represent signals at various points in the system, and the spectra on the $*$ and $\times$ lines, respectively, represent *spectral-convolution* and point-by-point *spectral multiplication* operations that relate those signals.

The first line in the figure depicts schematically a real IF signal with one- and two-sided bandwidths of 40 MHz and 80 MHz, respectively, and with positive- and negative-frequency components, respectively centered at 75 MHz and −75 MHz. The second line of Figure 25.8 shifts the IF signal by LO frequency −75 MHz using spectral convolution. (We'll see shortly how this is done in hardware.) The result, on the third line, has spectral components centered at 0 MHz and −150 MHz. Multiplication by the lowpass-filter frequency response in line 4 then removes the latter component, leaving only the complex baseband signal of line 5, which has a two-sided bandwidth and a Nyquist frequency of 40 MHz. The spectral convolution on line 6 corresponds

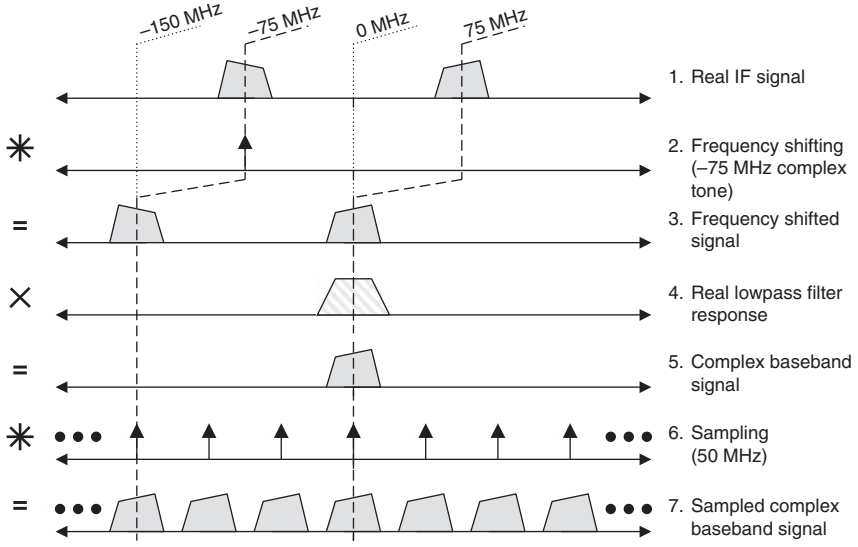


FIGURE 25.8 Analog downconversion in the frequency domain

to time-domain multiplication of a uniform impulse train at the 50 MHz sampling frequency with the signal represented by line 5. In the time domain, the result is a 50 MHz train of sampling impulses with areas that are (give or take a scale factor that we are ignoring) samples of the line 5 signal at the sampling instants. Of course, we will not create the line 7 impulses in hardware but will instead realize the impulse areas digitally as numbers in registers.

A block diagram showing how this process might be implemented in hardware is shown in Figure 25.9. The IF signal is sent to two mixers. In one mixer, the IF signal is beat with the 75 MHz LO with cosine phasing, and in the other mixer, the IF is beat with the same LO but with negative sine phasing, so that the mixers are operated in quadrature, 90° apart. The mixer outputs taken together as a complex pair form a complex signal with the spectrum shown in line 3 of the previous figure. These signals are then passed through lowpass filters (LPF) to remove the spectral component centered at −150 MHz that would otherwise have resulted in aliasing in the sampling step to follow.

Labels I (in-phase) and Q (quadrature) are traditionally used to indicate the real and imaginary parts of complex time-domain signals, like those here, that are realized as pairs of real signals. When a vertical cut through a diagram, such as in Figure 25.9, picks up one I signal and one Q signal, the represented complex signal crossing that cut is I + jQ. In the diagram, cuts just before and after the LPF blocks pick up complex signals with the spectra shown on lines 3 and 5 of Figure 25.8, respectively. The line 3 signal is created in the time domain as

$$\begin{aligned}[\text{line 3}] &= [\text{line 1}]\, e^{-j2\pi f_{\text{LO}}t} \\ &= [\text{line 1}]\cos(2\pi f_{\text{LO}}t) \\ &\quad -j[\text{line 1}]\sin(2\pi f_{\text{LO}}t) \\ &= [\text{I}_3 \\ &\quad + jQ_3]\end{aligned}$$

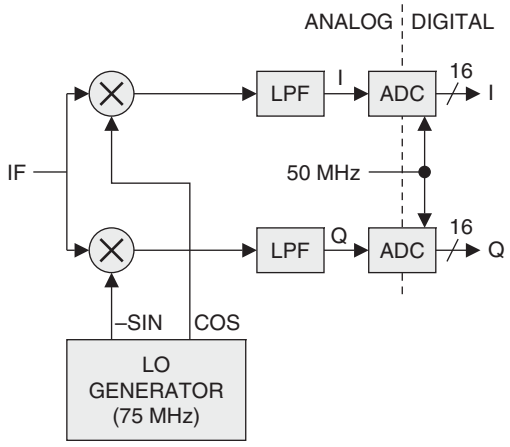


FIGURE 25.9 Typical analog downconversion to baseband and digitizer

where $f_{LO} = 75$ MHz. Similarly, using $*$ for time-domain convolution (filtering with an impulse response),

$$\begin{aligned}[\text{line 5}] &= [\text{line 3}] * h(t) \\ &= [\mathrm{I}_3 \\ &\quad + j\mathrm{Q}_3] * h(t) \\ &= (\mathrm{I}_3 * h(t)) \\ &\quad + j(\mathrm{Q}_3 * h(t)) \\ &= \mathrm{I}_5 \\ &\quad + j\mathrm{Q}_5\end{aligned}$$

When the filter outputs are viewed as complex signal [line 5] $= \mathrm{I}_5 + j\mathrm{Q}_5 = Ae^{j\theta}$, the complex magnitude A and angle θ give the amplitude (give or take a scale factor) and phase of the IF signal, because the original IF signal could be re-created from the line 5 signal in the time domain (again give or take a scale factor) as [line 1] $= \mathrm{Re}\{[\text{line 5}]\ e^{j2\pi f_{LO}t}\}$, from which it follows that, [line 1] $= \mathrm{Re}\{\mathrm{A}\ e^{j\theta}\ e^{j2\pi f_{LO}t}\} = \mathrm{A}\ \mathrm{Re}\{e^{j(2\pi f_{LO}t + \theta)}\}$ $= \mathrm{A}\cos(2\pi f_{LO}t + \theta)$.

As a final step the baseband I and Q signals after the filters are digitized by ADCs at a 50 MHz sampling rate, producing I_7 and Q_7 output samples or, equivalently, complex output samples $\mathrm{I}_7 + j\mathrm{Q}_7$.

The slash through the output of the ADC with a "16" above it in Figure 25.9 indicates that our ADC produces 16 bits of digital output. ADCs provide approximately 6 dB of dynamic range per bit, so our 16-bit ADC provides about 96 dB of dynamic range, assuming ADC nonlinearities are negligible.

A General Approach to Digital Downconversion. In digital downconversion, the analog IF signal is first sampled by an ADC, and all of the subsequent processing is then done digitally. Figure 25.10 depicts the digital downconversion process for our previous example, again in the frequency domain. The top line schematically represents the real IF signal with parameters as before. Performing the sampling analysis described previously, we discover that setting the sample rate to the two-sided signal bandwidth of 80 MHz would produce aliasing. However, a 100 MHz sample rate does not cause aliasing and is used on the second line of the figure. Sampling the input signal at 100 MHz replicates the signal spectrum at 100 MHz intervals as shown on line 3. Frequency shifting is accomplished by spectrally convolving this signal with the complex –75 MHz LO tone shown in line 4, producing the frequency-shifted signal on line 5. The latter signal is spectrally multiplied by the filter response shown on line 6 to remove the copies of the negative-frequency signal component, producing the complex baseband signal shown on line 7. This signal, which now has a two-sided bandwidth and Nyquist frequency of 40 MHz, is spectrally convolved in line 8 with impulses at the spectral origin and at 50 MHz to effectively decimate the signal by a factor of two.[3] The final baseband signal on line 9 has a sample rate of 50 MHz.

Figure 25.11 depicts the hardware implementation of this DDC architecture. The IF signal centered at 75 MHz is digitized directly by an ADC. After the ADC, the architecture is very similar to analog downconversion, except that the processing is performed digitally. In our example, we elect to sample the IF signal at a rate of 100 MHz with a 16-bit ADC. This architecture realizes the LO with a numerically controlled oscillator (NCO) that generates digital words to represent the cosine and negative sine signals at the LO frequency, here 75 MHz and sampled at the ADC

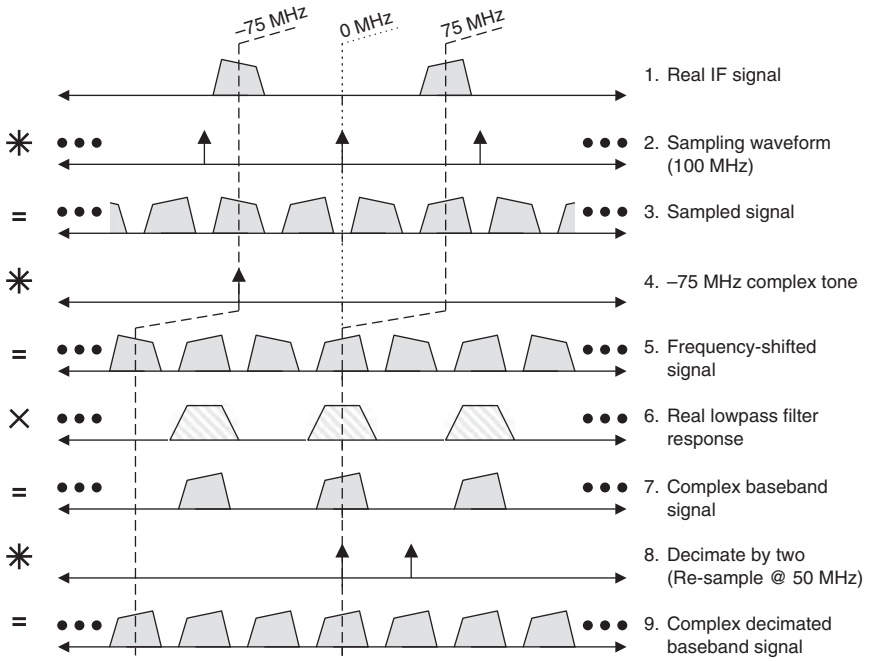

–75 MHz
0 MHz
75 MHz
1. Real IF signal
2. Sampling waveform (100 MHz)
3. Sampled signal
4. –75 MHz complex tone
5. Frequency-shifted signal
6. Real lowpass filter response
7. Complex baseband signal
8. Decimate by two (Re-sample @ 50 MHz)
9. Complex decimated baseband signal

FIGURE 25.10 Digital downconversion in the frequency domain

sample rate. The sine and cosine signals from the NCO are then digitally multiplied by the digitized IF signal. In this particular example, the relationship between the LO frequency and the sampling rate will make the required NCO and the multipliers both rather trivial because each required NCO output value is zero or ±1, and that special case will be addressed shortly. For now, this architecture supposes that no such special situation exists and that a general NCO/multiplier structure is needed. The design of

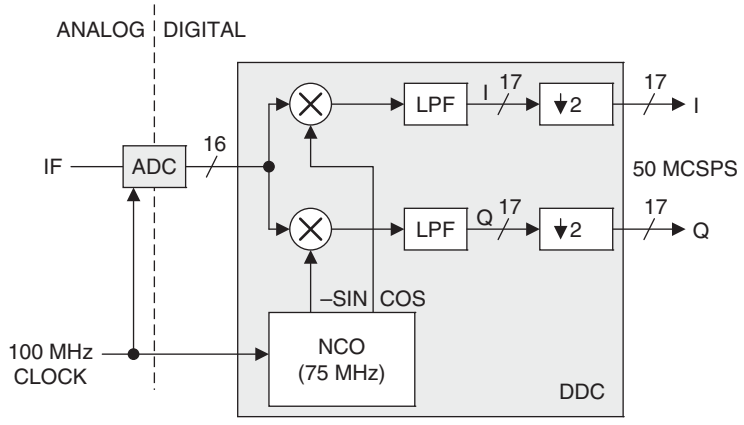


FIGURE 25.11 Digital downconversion architecture

a general NCO will be in Section 25.3. Following the multiplications, digital lowpass filters prevent aliasing when their outputs are decimated by a factor of two to produce complex output samples at a 50 MHz rate. In the figure, MCSPS stands for *million complex samples per second*.

The lowpass filter also reduces out-of-band noise and thus increases signal-to-noise ratio (SNR). In order to preserve this SNR growth, the number of bits used to represent the filter output might need to increase. If the filter reduces the bandwidth of the data by a factor R without affecting the signal of interest, then the SNR increase in dB is given by $10\log_{10}R$. In our example, a 2× reduction in bandwidth results in approximately a 3 dB increase in SNR. With each bit representing about 6 dB of SNR, the minimum number of bits required to represent the filtered signal could grow from 16 to 17.

In an actual application, the system designer needs to analyze the effects of sampling and digital processing and determine how many bits need to be carried through the calculations in order to preserve SNR and prevent overflow. Considerations include front-end system noise, which is typically allowed to toggle two or more bits (four or more quantization levels) of the ADC output. Also, an actual N-bit ADC never provides exactly 6N dB of SNR, due to ADC-induced errors. For example, a typical 16-bit ADC generally provides about 14 bits or about 84 dB of SNR. Carrying 16 bits through the signal processing provides about 96 dB of dynamic range. In this case, the designer may elect to allow the datapath through the lowpass filter to remain at 16 bits, realizing that the filtering process has simply increased the SNR of the signal to 87 dB, which could still be accommodated by the 16-bit datapath.

A DDC provides several benefits compared to analog downconversion. The analog approach is subject to various hardware errors, including mismatch of the mixers, LO signals not exactly 90° apart, and mismatches in the gains, DC offsets, or frequency responses of the I and Q signal paths. A DDC avoids these problems, though it is vulnerable to the phase noise of the ADC sample clock, ADC nonlinearities, and arithmetic round-off noise. Realizing maximum performance requires careful attention to design details.

Direct Digital Downconversion. If the designer has some flexibility in either the IF center frequency or ADC sample rate, a simplified DDC architecture, direct digital downconversion, can be considered.[4,5] If the ADC sample rate is four times the center of the IF band, then the sampling process can also shift the spectrum to baseband, and the NCO and associated multipliers of the general DDC are not needed. In general, direct conversion to baseband is a simple and cost-effective DDC method that can be used when the signal being sampled is always centered at a single frequency. The standard DDC architecture might need to be used when the center frequency of the signal being sampled dynamically changes, which forces the DDC's LO to change accordingly.

Let's look at the direct DDC in the time domain first, for intuition, and then we can carefully derive the architecture in the frequency domain. Suppose the DDC architecture is as sketched in Figure 25.11, with an IF centered at 75 MHz and a 75 MHz LO and suppose the NCO is set to 300 MHz so that it produces the sampled sines and cosines shown in Figure 25.12*a*, where vertical lines and dots indicate sample times and values, respectively. Because the sample rate is four times the LO frequency, the (cos, -sin) LO sample pairs cycle repeatedly through (1, 0), (0,–1), (–1, 0), and (0, 1).

Next, suppose the IF signal is a 75 MHz sinusoid of arbitrary phase as in line (b). The DDC's mixer outputs I and Q, the products of the line (b) IF signal with the two

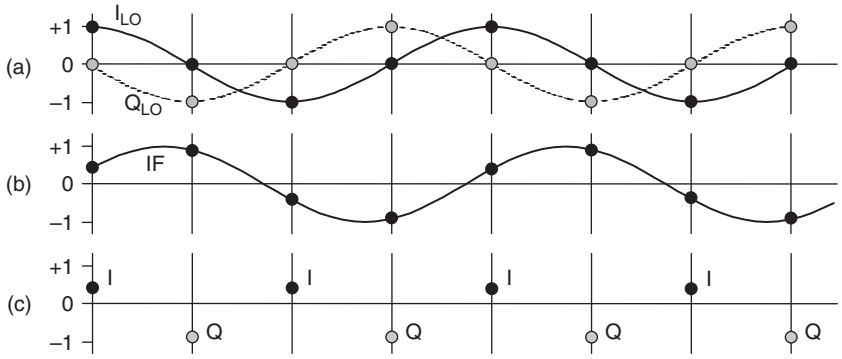


FIGURE 25.12 Various signals sampled at 300 MHz: (*a*) 75 MHz cosine and –sine LO signals, (*b*) 75 MHz IF tone, and (c) result of multiplying (*a*) samples by (*b*) samples

line (a) LO signals, are then as in line (c). Because our hypothetical IF signal on line (b) was exactly at one quarter of the sample rate, both I and Q are constants, the sine and cosine of the IF signal's phase angle.

Figure 25.13 shows the same 75 MHz IF tone, but sampled at 100 MHz and 60 MHz, which are odd integer submultiples (1/3 and 1/5) of the original sample rate of 4× the IF center frequency, 300 MHz. Note that odd samples still cycle between I and –I, and the even samples switch between Q and –Q. Odd integer submultiples of 4× the IF center frequency can, therefore, be viable alternative sample rates. A Nyquist bound applies and requires the two-sided IF bandwidth to be less than the sampling rate.

Now let's derive the direct-DDC architecture carefully in the frequency domain. Suppose a real IF signal is once again centered at 75 MHz and sampled at 100 MHz as in line (a) of Figure 25.13. The first three lines of Figure 25.14 illustrate this in the frequency domain with line 3 showing the sampled IF signal. The bandpass filter response on line 4 removes the unwanted spectral components to produce the complex passband signal of line 5. This signal is then decimated by 2 and shifted by –75 MHz to produce, at a 50 MHz sampling rate, the desired complex baseband signal shown on line 9.

Figure 25.15 shows the corresponding block diagram. The magnitude of the frequency response on line 4 of Figure 25.14 is neither an even nor an odd function, so the corresponding impulse response is neither purely real nor purely imaginary.

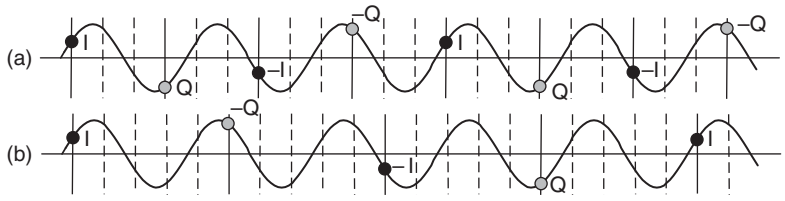


FIGURE 25.13 75 MHz tone sampled at (*a*) 100 MHz (4/3 × IF) and (*b*) 60 MHz (4/5 × IF)

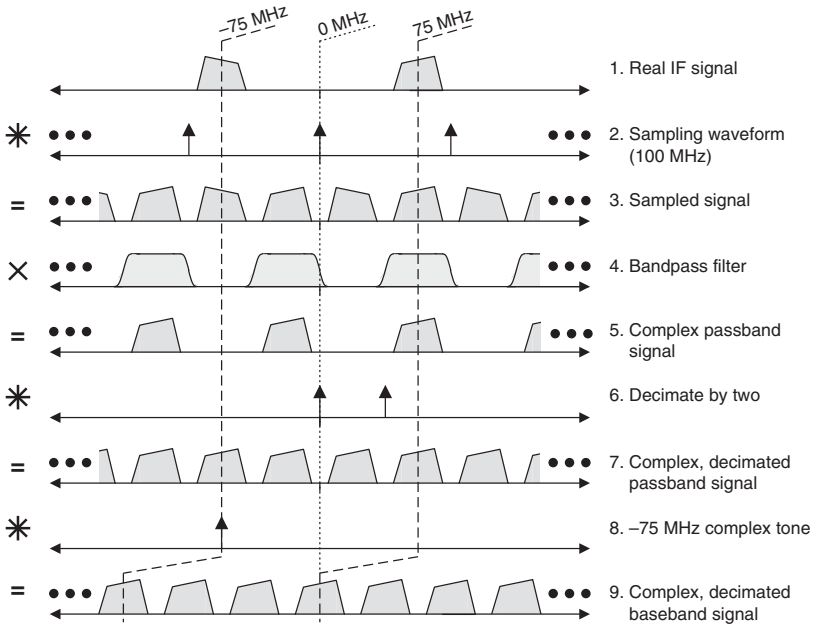


FIGURE 25.14 Direct digital downconversion in the frequency domain

Writing that impulse response as $h(n) = h_I(n) + jh_Q(n)$, using real functions $h_I(n)$ and $h_Q(n)$, the line 4 operation becomes

$$\begin{aligned}[\text{line 5}] &= [\text{line 3}] * h(n) \\ &= [\text{line 3}] * [h_I(n) \\ &\quad + jh_Q(n)] \\ &= ([\text{line 3}] * h_I(n)) \\ &\quad + j([\text{line 3}] * jh_Q(n)) \\ &= I_5 \\ &\quad + jQ_5\end{aligned}$$

where the fact that line 3 is real in the time domain was used in the last step. In Figure 25.15, the sampled IF signal, therefore, passes through different FIR filters

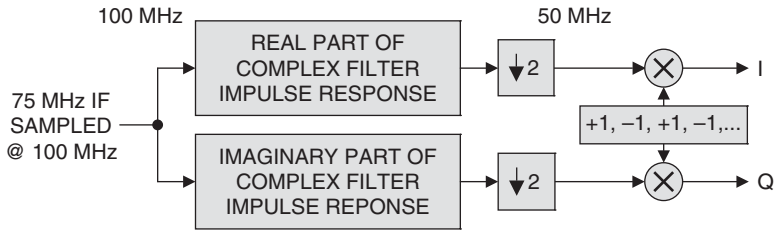


FIGURE 25.15 Time-domain implementation of a direct digital downconverter

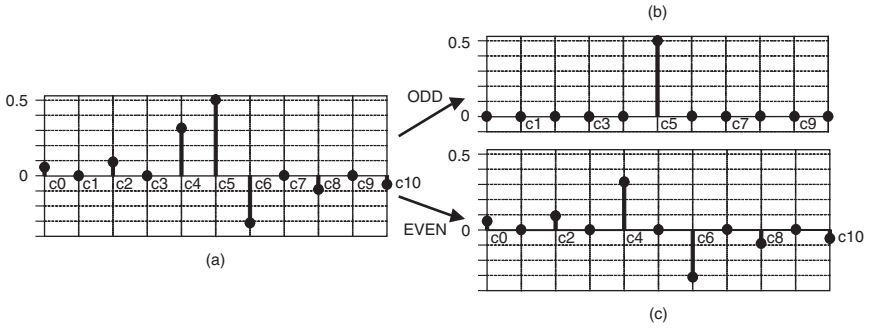


FIGURE 25.16 (*a*) Halfband bandpass filter coefficients for a direct digital downconverter, (*b*) real (odd), and (*c*) imaginary (even) parts of the complex impulse response

(FIR filters are described in Section 25.4), where the top and bottom filters apply the real and imaginary parts of the coefficients, respectively. The equivalent complex-impulse-response filter, with the frequency response shown in line 4 of Figure 25.14, is a halfband filter because that frequency response and a version shifted in frequency by half a period sum to a constant. This property causes almost half of its impulse-response coefficients to be zero. Figure 25.16*a* illustrates the coefficients of a typical filter for this application. All of the odd-numbered coefficients, except for the one in the center, are zero, so the filter is very efficient to implement, as the zero coefficients don't require multipliers. The frequency response's symmetry about 1/4 of the sampling rate causes the even-and odd-numbered coefficients to be purely real and purely imaginary, respectively, so the even-and odd-numbered coefficients are used to, respectively, create I and Q, as shown in Figure 25.16*b* and *c*.

After the filters, the complex signal is decimated by 2 to produce a 50 MHz output sample rate. The final spectral convolution by a –75 MHz tone is accomplished by negating every other sample.

In Figure 25.17, we transform the system of Figure 25.15 to make it more computationally efficient. We begin with the structure in Figure 25.17*a*, which shows the filtering in detail using τ to indicate each clock-interval delay. The location of the one nonzero coefficient in the real part $h_I(n)$ of the Figure 25.16 impulse response corresponds to an odd-numbered delay, so $h_I(n)$ is realized using a single delay and some number of double delays. The imaginary part $h_Q(n)$ of the impulse response, in contrast, has nonzero coefficients only at even numbers of delays, so it is realized with double delays only.

The architecture can then be further simplified by moving the decimation ahead of the 2τ delays, as shown in Figure 25.17*b*. This changes each double delay to a single delay at the lower clock rate at which the filter computations are now more efficiently clocked. Optionally, the negation of alternate samples at the output can now be relocated to the decimation's output. Each delay that the negation crosses as it moves in this transformation causes a net sign change in the signal, so each signal path between the old location and the new that contains odd numbers of delays requires coefficient negation to compensate. The result in the design of Figure 25.17*c* is negation of alternate coefficients in the Q filter, as shown in Figure 25.17*c*.

The optional negation-moving transformation just described yields a simple interpretation of system operation. Figure 25.12 shows that the leading τ delay, decimation, and sign-negation operations of Figure 25.17*c* work together to steer I and Q samples into the upper and lower filter paths, respectively, but the samples that are

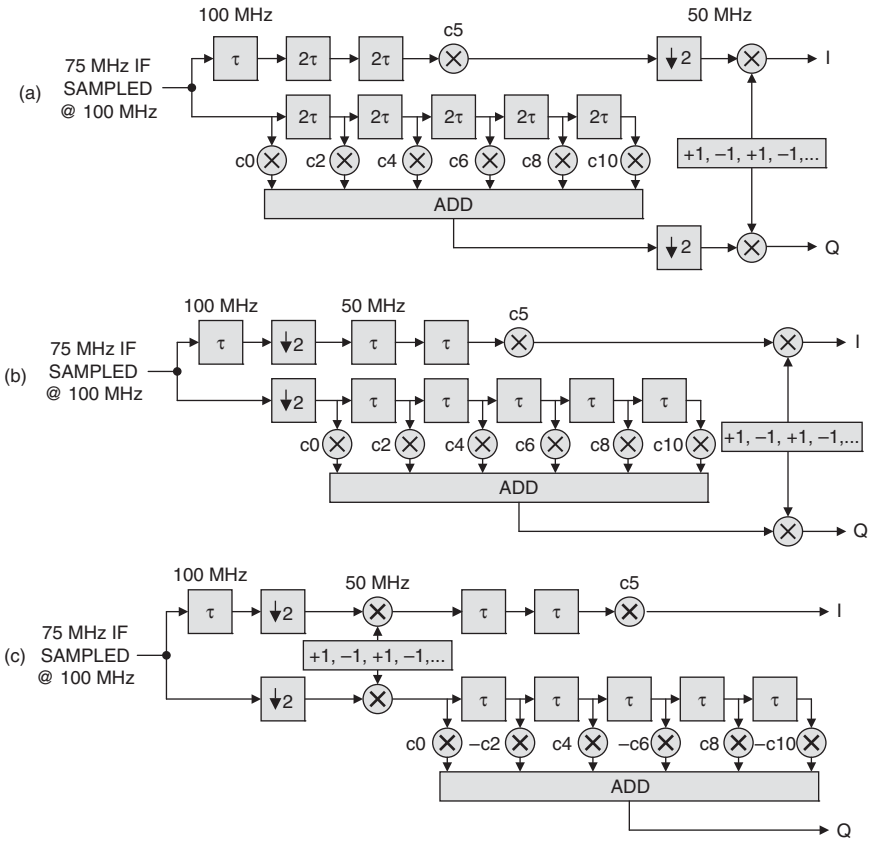


FIGURE 25.17 Direct digital downconverter: (*a*) baseline implementation, (*b*) decimating before filters, and (*c*) inverting every other sample after decimation

now aligned in time as they pass through the remaining processing do not actually correspond to the same points on the IF signal input's time line, since the I and Q samples were derived from alternate ADC samples. However, the Q filter with alternate coefficients negated, shown in Figure 25.18, actually approximates the half-sample delay

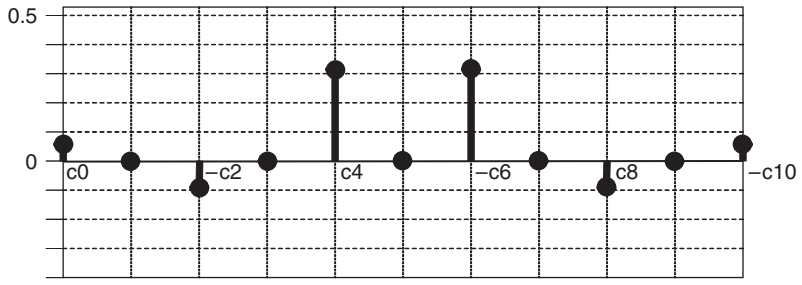


FIGURE 25.18 Negated-alternate-signs version of Q filter coefficients

required to realign the data in the two paths, and this causes the I and Q output values to be effectively sampled at the same instants.

Signal-Sampling Considerations. Actual devices and signals introduce errors. For example, clock jitter results in errors in the sampled output of an ADC, as shown in Figure 25.19. In addition, real ADCs also add internal jitter, or *aperture uncertainty*, which must be taken into account.[6] If the errors in the effective sampling instant introduced by these jitters are uncorrelated, a reasonable approximation, the RMS sample-time jitter they introduce, t_J, is

$$t_J = \sqrt{\left[(t_{J(\mathrm{ADC})})^2 + (t_{J(\mathrm{CLOCK})})^2\right]}$$

where $t_{J(\mathrm{ADC})}$ and $t_{J(\mathrm{CLOCK})}$ are the RMS sample time jitters introduced by the ADC and the clock, respectively.

A sinusoidal input signal of amplitude A and frequency f is expressed as

$$v(t) = A\sin(2\pi ft)$$

which has derivative

$$dv(t)/dt = A2\pi f\cos(2\pi ft)$$

The maximum error due to jitter occurs at $t = 0$, when the derivative of the signal is at its peak, or

$$dv(0)/dt = A2\pi f$$

The RMS error voltage, V_e, produced by an RMS sample time jitter, t_J, is given by

$$V_e = A2\pi f\, t_J$$

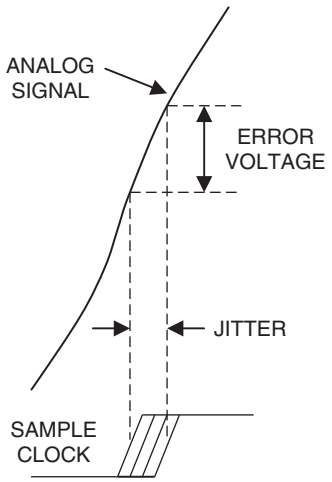


FIGURE 25.19 RMS jitter vs. RMS noise

This error voltage limits the theoretical maximum SNR of an ADC by

$$\mathrm{SNR}_{\max} = 20\log(A/V_e) = -20\log[(2\pi f t_J]$$

This relationship is presented in Figure 25.20, which plots on the left and right axes, respectively, the SNR and the equivalent ADC effective number of bits or ENOB (≈ SNR/6 dB), both versus analog frequency and for different values of RMS sample jitter. Due to a variety of error sources internal to an ADC (aperture uncertainty, nonlinearities, added noise, etc.), the specified ENOB of an ADC is always less than the number of bits it provides. For example, a 14-bit ADC typically has an ENOB of 12.

With the bandpass sampling technique described earlier, where the ADC can sample at a rate that is considerably lower than the analog frequencies being sampled, it might seem attractive to do away with the receiver altogether and sample the RF signal directly. Although this is possible, ADC limitations restrict the performance of such architectures. First, the analog front end of an ADC has a lowpass 3 dB cutoff frequency specified by the manufacturer. ADC input frequencies should be kept well below this cutoff. Second, as shown previously in Figure 25.20, sampling the RF signal directly will dramatically increase the slew rate of the signal presented to the ADC, thus requiring very low levels of RMS clock jitter. Also, the ADC has inherent nonlinearities that produce spurs in the ADC output, a problem which typically worsens with increasing input frequency. ADC datasheets specify the *spur-free dynamic range (SFDR)* of the device, which is typically defined as the dB difference in signal level between the desired signal and the largest spur measured at the ADC output with a single tone applied to the input. The SFDR of a typical ADC is higher than its specified SNR. Unfortunately, there are many definitions of SFDR, so the designer is advised to read manufacturers' datasheets carefully in this regard. As mentioned earlier, the SNR of a sampled signal can be increased by filtering to eliminate noise in parts of the spectrum that are otherwise unused. However, the spurs generated by an ADC may lie in the band of interest, where filtering would be inappropriate. Therefore, spurs lower than the unfiltered noise level can become relatively significant after ADC noise is reduced through filtering.

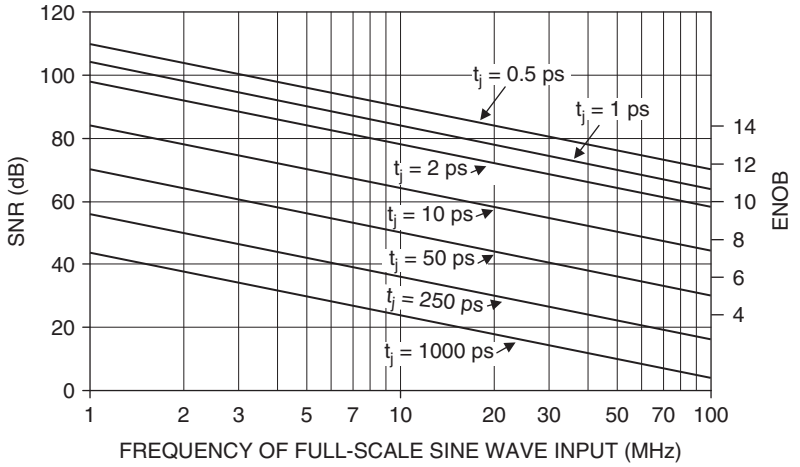


FIGURE 25.20 Signal-to-noise ratio vs. analog frequency for varying sample jitter

Multi-Beam Digital Beamforming. An important application of digital technology is for the beamforming function in a phased array antenna system. Figure 25.21*a* depicts an analog beamforming system. The wavefront shown can be thought of as the return from a target of interest. Note that the wavefront will hit each element of the array at different times. In order to form a beam in that particular direction, each element of the array needs to be followed by a time delay unit that delays the signal received at each element by the appropriate amount, such that when all of the outputs of the time delays are summed, they add up coherently to form a beam in the desired direction. If the system has a narrow bandwidth (bandwidth < ~5% of RF frequency) and the antenna beamwidth is not too narrow (so that the 3 dB beamwidth in degrees is greater than the percent bandwidth), the time delay can be approximated using phase shifters. Wide bandwidth systems require "true" time delays in order to form the beams and preserve the bandwidth. The receiver would follow the analog beamformer, as shown in the figure. Figure 25.21*b* shows an extreme application of digital beamforming, where

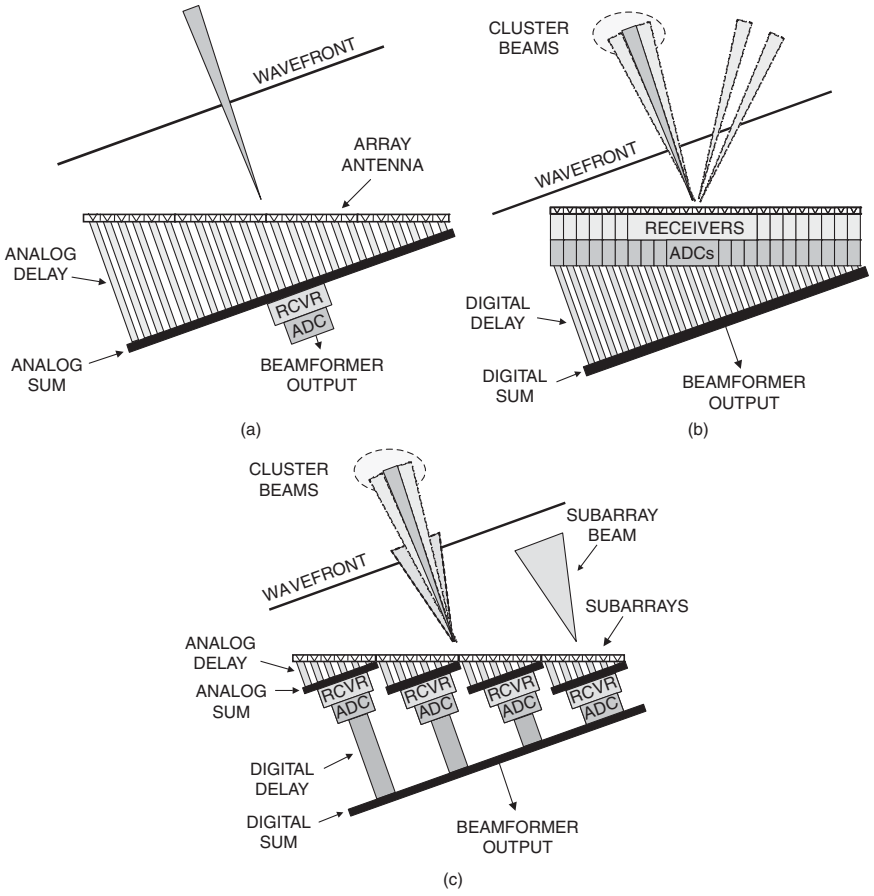


FIGURE 25.21 (*a*) Analog beamformer, (*b*) every-element digital beamformer, and (*c*) subarray digital beamformer

a receiver and ADC are behind every element. In this system, the time delay is implemented either as a digital phase shift or digital time delay, followed by a digital summer. This configuration allows beams to be formed in any direction, and multiple beams can be formed simultaneously, if desired, by using the same sample data and implementing different time delays to form the different beams. However, at this writing, putting a digital receiver behind every element is expensive and is usually not feasible for most large antenna applications (i.e., for systems with thousands of elements). One compromise solution is shown in Figure 25.21*c*, where analog beamforming is used to implement subarrays, which are followed by digital receivers and digital time delays.

Digital beamforming offers several advantages over analog beamforming. With an analog beamformer, usually only one beam is formed at a time. Radars are typically required to perform multiple functions, such as volume surveillance, target confirmation, tracking, etc. With only one beam at a time, there may not be enough time available to perform all of the required functions. A digital beamformer allows the formation of multiple simultaneous beams, allowing the volume surveillance function to be performed much more quickly, allowing more time to do other things. Of course, in order to form multiple simultaneous receive beams, the transmitted beam must be made broader to encompass the receive beams, which might require a more powerful transmitter or more integration on receive to provide the same performance as a single-beam system.

Another advantage has to do with dynamic range. In an analog beamforming system, there is only one receiver and ADC, and the dynamic range performance is limited to the capability of a single channel. In a digital beamforming system, there are multiple receivers and ADCs, and the number of ADCs that are combined determines the system dynamic range. For example, if the outputs of 100 ADCs were combined to form a beam, assuming that each ADC induces noise that is of equal amplitude and uncorrelated with the others, there would be a 20 dB increase in system dynamic range, compared to a single-receiver system using the same ADC.

Figure 25.22 shows a block diagram of a typical digital beamforming system. Each antenna output port, be it from an element or a subarray, is followed by a digital

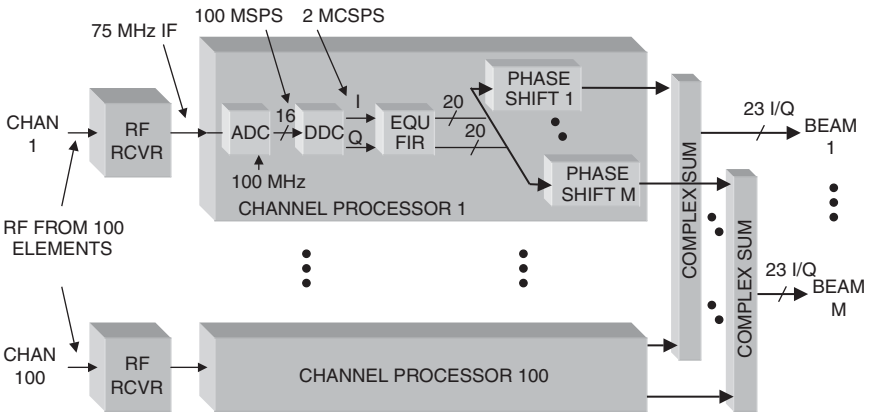


FIGURE 25.22 Typical digital beamformer

downconverter and an equalization filter (EQU FIR). The equalization filter is typically a complex finite impulse response (FIR) filter (described later) that adjusts the frequency response of each channel so that its passband matches the other channels in phase and amplitude before it is summed with the other channels in the beamformer. The coefficients of this filter are determined through a calibration process. During calibration, a test signal is presented to the RF input of all channels. This signal is typically a swept frequency tone or a noise input that covers the channel bandwidth. The ADC samples of all channels are collected simultaneously and complex weights are calculated for the equalization filters that force the frequency response of each channel to be matched. Once the channel is equalized, a unique time delay is implemented for each beam to be formed. As mentioned earlier, this time delay can be realized either as a phase shift for a narrowband system or as a time delay for a wideband system. A phase shift can be implemented with a complex multiply or a CORDIC operation, both of which will be described later. A time delay can be implemented with a FIR filter that imposes a linearly changing phase shift over frequency on the signal. Once the time delay is realized in each channel, the appropriate complex time-delayed signals from all of the channels are summed to form a beam. M complex summers are required to form M beams.

Digital Pulse Compression. Pulse compression is another signal processing function that is predominantly being performed digitally in radar systems. However, at this writing, many systems still exist with analog-delay-line pulse compressors. In these systems, analog pulse compression is performed at an IF, followed by the ADC in the processing chain. Because pulse compression increases the SNR of the signal, performing it before sampling increases the dynamic range requirement of the ADC. In a digital pulse compression system, the ADC precedes the pulse compressor and only has to accommodate the precompression dynamic range of the signal, which can be a significantly lower requirement. The digitized signal is converted to baseband and passed to the digital pulse compressor. The increased dynamic range due to the pulse compression gain is accommodated by increasing the number of bits in the digital computation.

Chapter 8 is devoted totally to pulse compression radar. In summary, there are two basic approaches to implementing digital pulse compression: time-domain and frequency-domain convolution. A generic time-domain convolver consists of a complex FIR filter, where the coefficients are the complex conjugate of the transmitted baseband waveform samples in time-reversed order (which is also the definition of the matched filter for the transmitted signal). This architecture can compress any arbitrary waveform. A simplified version of the architecture results when the modulation is a binary phase code. In this case, the coefficients are either +1 or –1, so the arithmetic performed for each sample is a complex sum or subtraction instead of a full complex multiplication.

Pulse compression may also be accomplished by operating in the frequency domain, where it is referred to as *fast convolution.* In this case, the baseband samples of the receive data and the reference transmit waveform are passed through fast Fourier transforms (FFTs), the data FFT outputs are multiplied point-by-point by the complex conjugate of the reference FFT outputs, and then the result is converted back to the time domain by an inverse FFT. In general, it is more hardware efficient to perform time-domain convolution for a small number of coefficients and frequency domain convolution for a large number (more than 8 or 16) coefficients.

25.3 TRANSMIT CHANNEL PROCESSING

Before digital technology became widely available, analog techniques were employed to generate radar transmit waveforms. Simple pulsed systems used analog RF switches to gate the LO on and off. Frequency modulated signals were generated by surface acoustic wave (SAW) devices. Simple binary phase modulation schemes like pseudo-random noise waveforms were also possible. Digital technology, however, presents the radar system designer with many more options and allows arbitrarily modulated transmit waveforms to be modified pulse-to-pulse if desired. This section describes several of the techniques commonly used to generate the radar transmit signal digitally.

Direct Digital Synthesizer (DDS). Figure 25.23 shows a block diagram of this technique, in which a numerically controlled oscillator (NCO) generates a digitized sinusoid that is converted to an analog signal by a digital-to-analog converter (DAC). Figure 25.24 demonstrates how an NCO operates to produce a sine wave. The n-bit tuning word is actually a phase increment that determines the frequency of the sine wave output. The phase increment is expressed in a format called Binary Angle Measurement (BAM), in which the most significant bit (MSB) of the word represents 180°, the next bit represents 90°, and so on. In the phase accumulator, the tuning word is added to the output of a running sum, implemented as an adder followed by a register (REG). This produces a uniformly increasing phase, incremented at the system clock rate. The m MSBs of the running sum are sent to a phase-to-amplitude converter, which is a lookup table that produces a k-bit value that represents the amplitude of the sine wave at the input phase. If we represent the tuning word by M, the sample frequency by f_s, and the number of bits in the phase accumulator by n, then the frequency of the output sine wave can be expressed as $Mf_s/2^n$.

In this scheme, the phase represented by the running sum will overflow when it crosses over 360°. The advantage of expressing the phase in BAM notation is that it allows modulo-2π arithmetic and overflows are automatically taken care of, since a 360° phase shift is the same as 0°. For example, assume we have a 3-bit BAM notation, which means that the least significant bit (LSB) represents a phase shift of 45°. Let's also assume that the tuning word is represented by 001 for a phase increment of 45° every clock. The running sum phase would steadily increase on every clock edge, becoming 000 (0°), 001 (45°), …, 110 (270°), and 111 (315°). On the next clock edge, the phase should be represented by 1000 for 360°. However, we are only provided a 3-bit adder, so the MSB is simply lost, leaving us with a phase code of 000 (0°), which is the same as 360°. Therefore, the resulting phase waveform takes on a sawtooth pattern, linearly ramping from 0° to not quite 360° and then resetting to 0° and ramping again.

An important feature of an NCO for a radar application is the CLEAR signal shown going to the phase accumulator register. For a coherent radar exciter implementation, the transmit signal must start at the same phase on every pulse. Otherwise, the

FIGURE 25.23 Direct Digital Synthesizer (DDS)

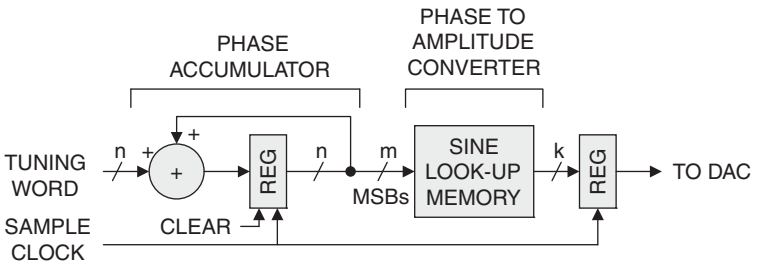


FIGURE 25.24 NCO block diagram

transmitted signal will have arbitrary phase from pulse to pulse, making doppler processing difficult if not impossible. The CLEAR control provides the means to do this. In some applications, like for a transmit beamformer, the starting phase in each channel may need to be different in order to steer the beam. In this case, we could provide a mechanism to set the phase to the desired value at the beginning of a pulse instead of simply clearing it.

The DDS can also be used to generate linear and nonlinear FM "chirp" waveforms. This is accomplished by providing circuitry that changes the tuning word from sample to sample in order to provide the desired frequency (or phase) modulation. For example, a linear FM chirp waveform requires a phase that changes in a square-law fashion with time. This can be accomplished by changing the tuning word (or phase step size) in a linearly increasing or decreasing way on every sample.

Digital Upconverter (DUC). Another popular method to implement a transmit waveform is through digital upconversion, also referred to as *arbitrary waveform generation.* In this technique, a digital complex baseband waveform, usually read from a memory, is first interpolated to a higher sample rate, and then modulated with digitized sine and cosine signals to produce a modulated carrier. Figure 25.25 provides a block diagram of a DUC that translates a complex baseband signal up to a 25 MHz IF. The baseband I and Q signals enter the DUC at a rate of 2 MCSPS and are first up-sampled by a factor of 50. This is accomplished by inserting 49 zeroes between each input sample and increasing the clock rate to 100 MHz. This signal is then passed through a digital lowpass filter that performs the interpolation. These signals are then multiplied

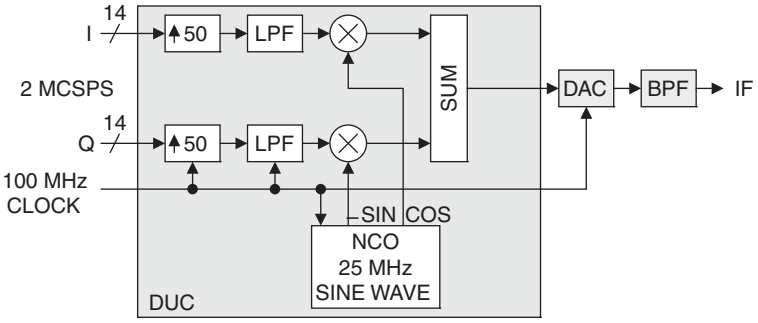


FIGURE 25.25 Digital upconverter (DUC)

by digitized sine and cosine waveforms for the modulation carrier frequency, producing a complex modulated IF as an output. These signals are digitally summed, converted to analog through a DAC, and passed through a bandpass filter to produce an IF output. For large upsampling ratios, a cascaded-integrator comb (CIC) interpolator, described in the next section, provides an efficient implementation.

25.4 DSP TOOLS

This section will describe various processing architectures and techniques that are available to DSP engineers.

Phase Shift. The phase shift is a core element in DSP design, and there are several techniques available to implement one. The most straightforward approach is to simply perform a complex multiply, as shown in Figure 25.26. In this example, the complex input sample is denoted as $A + jB$, which is multiplied by the complex coefficient $C + jD$ to produce $(AC - BD) + j(AD + BC)$ in order to effect the phase shift. This operation requires four multipliers and two adders.

After some manipulation, the following can be shown:

$$I = (AC - BD) = D(A - B) + A(C - D)$$

$$Q = (AD + BC) = C(A + B) - A(C - D)$$

Noting that the final term is the same in both equations, we see that this complex multiplier can be implemented with only three real multipliers and five real adds. This can be important if real multipliers are at a premium. Figure 25.27 shows a block diagram of this architecture.

CORDIC Processor. An efficient and versatile algorithm that can implement a phase shift without using multipliers is the **CO**ordinate **R**otation **DI**gital **C**omputer (CORDIC) function, first described by Volder[7] in 1959. The CORDIC can implement

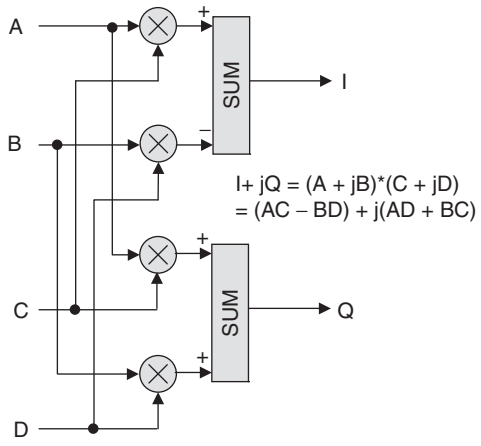


FIGURE 25.26 Standard complex multiply

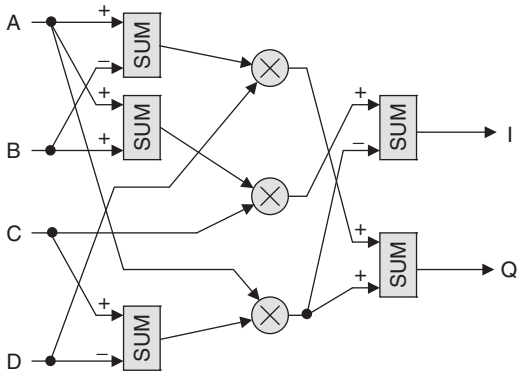


FIGURE 25.27 Complex multiply with three real multipliers

a variety of functions, including sine, cosine, vector rotation (phase shift), polar-to-rectangular and rectangular-to-polar conversions, arctangent, arcsine, arccosine, and vector magnitude, through an iterative process that just uses bit shifts and adds.[8] The following discussion describes the CORDIC algorithm.

The equations that shift the phase of complex number $I_0 + jQ_0$ by an angle θ to produce $I_1 + jQ_1$ are as follows:

$$I_1 = I_0(\cos(\theta)) - Q_0(\sin(\theta))$$

$$Q_1 = I_0(\sin(\theta)) + Q_0(\cos(\theta))$$

These equations can be manipulated to provide

$$I_1 = \cos(\theta)[I_0 - Q_0(\tan(\theta))]$$
$$Q_1 = \cos(\theta)[Q_0 + I_0(\tan(\theta))]$$

The CORDIC algorithm takes advantage of this relationship to approximate an arbitrary phase shift by implementing multiple stages of phase shifts, where the tangent of the phase shift in each successive stage is the next smaller fractional power of 2, and multiplication by this number can be implemented by shifting the input data bits an integer number of places. The first few stages are as follows:

$$I_1 = \cos(\theta_0)[I_0 - Q_0(\tan(\theta_0))] = \cos(\theta_0)[I_0 - Q_0(1)]$$
$$Q_1 = \cos(\theta_0)[Q_0 + I_0(\tan(\theta_0))] = \cos(\theta_0)[Q_0 + I_0(1)]$$

$$I_2 = \cos(\theta_1)[I_1 - Q_1(\tan(\theta_1))] = \cos(\theta_1)[I_1 - Q_1(½)]$$
$$Q_2 = \cos(\theta_1)[Q_1 + I_1(\tan(\theta_1))] = \cos(\theta_1)[Q_1 + I_1(½)]$$

Table 25.1 shows these parameters for an eight-stage CORDIC processor. Each row of the table represents successive iterations of the algorithm. The **tan(θ_i)** column shows the factor by which the I and Q values are multiplied for each iteration. Note that these values are fractional powers of 2, so the multiplication can be realized by shifting the binary I and Q values right by i places. The θ_i column shows the arctangent of this factor, which can also be thought of as the phase shift applied during each iteration.

TABLE 25.1 CORDIC Parameters for First Eight Stages

i	$\tan(\theta_i)$	θ_i (deg)	$\cos(\theta_i)$	P $[\cos(\theta_i)]$
0	1	45.000	0.707107	0.707107
1	1/2	26.565	0.894427	0.632456
2	1/4	14.036	0.970143	0.613572
3	1/8	7.1250	0.992278	0.608834
4	1/16	3.5763	0.998053	0.607648
5	1/32	1.7899	0.999512	0.607352
6	1/64	0.8951	0.999878	0.607278
7	1/128	0.4476	0.999970	0.607259

The $\cos(\theta_i)$ column shows the cosine of this angle, which should be multiplied by the result of each iteration, as shown in the equations above. In actual applications, however, this cosine multiplication is not performed at every iteration. At each stage, the implied factor that needs to be multiplied by the IQ outputs of the stage in order to provide the correct answer is the product of all of the cosines up to that point, as shown in the P**[cos(θ_i)]** column. For a large number of iterations, this product of cosines converges to a value of 0.607253. In most cases, this scaling can be compensated for in later stages of processing. The inverse of this factor, 1.64676 for a large number of iterations, is the processing gain imposed on the IQ results of the CORDIC. If integer arithmetic is performed, an extra bit should be provided at the most significant end of the adders in order to accommodate this increased signal level.

Figure 25.28 is a flow chart that represents the CORDIC algorithm to implement a phase shift. The inputs to the algorithm are the I_{in}, Q_{in}, and ϕ_{in} (the desired phase shift). The variable i will keep track of the processing stage being performed and is initialized to zero. The basic algorithm can perform a phase shift between ±90°. If the desired phase shift is outside of that range, the input I and Q values are first negated, imposing a 180° phase shift, and 180° is subtracted from the desired phase shift. The new phase shift is now within ±90°, and the algorithm proceeds normally.

Next, the algorithm loops through N iterations with the goal of driving the residual phase error, ϕ, to zero. In each iteration, a new ϕ is calculated by subtracting or adding the phase shift for

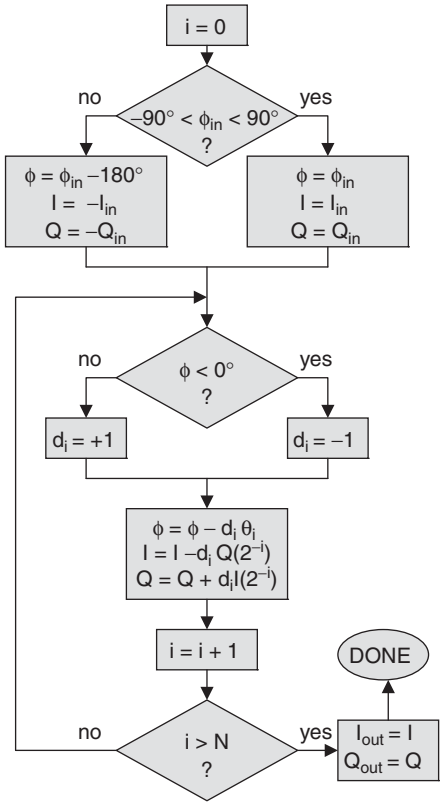


FIGURE. 25.28 CORDIC algorithm flow chart

that stage (θ_i from the table) to the previous value of ϕ. If $\phi < 0$, θ_i is added to ϕ. Otherwise, θ_i is subtracted from ϕ. In each stage, the Q (or I) input is divided by a factor of 2^i by shifting the number to the right by i bits, and then added to or subtracted from the I (or Q) input, depending on the sign of ϕ. The variable i is incremented and the process repeats until $i > N$, at which point the phase-shifted results are available.

Figure 25.29 is a block diagram of an eight-stage CORDIC processor that implements a phase shift, where each stage represents an iteration in the flow chart. An N-stage processor provides a phase shift that is accurate to within $\pm\theta_N$ degrees (from the table), so the more stages in the processor, the more accurate the answer. The input I and Q values change on the rising edge of an assumed sample clock. In the first stage, the I value is either added to or subtracted from the Q value in the ADD/SUB block. The control block on the bottom of the figure determines whether additions or subtractions are performed at each stage, based on the algorithm described previously. If the ADD/SUB block in the Q channel performs an addition, the same block in the I channel will perform a subtraction, and vice-versa. The result of the ADD/SUB blocks is stored in a register (REG) on the next clock edge and passed to the next stage of processing. In this implementation the last block labeled (PASS/INV) performs the required inversion of I and Q if the desired phase shift is beyond the $\pm 90^o$ range of the algorithm. The final multiplication by a constant is optional, as described earlier.

The architecture shown in Figure 25.29 is a good example of a *pipelined* processor, in which a portion of the computation is performed and the result is stored in a register on each rising edge of the sample clock and passed to the next stage of processing. The processor would still function if the registers were removed. However, in that case, when the input I and Q values changed, the final output would not be useable until the results of the new input values rippled through all of the stages of processing, which would generally be an unacceptably long period of time. In a pipelined processor, a small portion of the total calculations is performed at a time, and the result is stored in a register and passed to the next processing stage. This architecture provides a higher *throughput* than the nonpipelined version, which means that the final result can be produced at a much higher sample rate, which is inversely proportional to the delay of a single stage. The *latency* of a pipelined processor refers to the delay experienced between the time a new data sample is entered into the processor and the time that the result based on that input is available on the output. The eight-stage, pipelined CORDIC processor shown in the figure would have a latency equivalent to eight clock periods and a throughput equivalent to the clock rate (i.e., once the pipeline is filled and the first result is available on the output, successive clocks will produce new outputs at the clock rate).

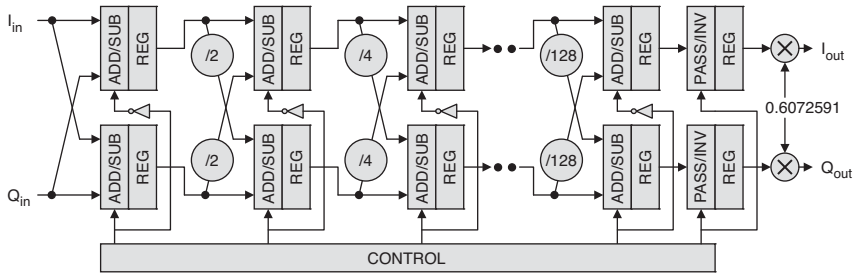


FIGURE 25.29 Eight-stage CORDIC processor

Digital Filters and Applications. This section describes several of the major forms of digital filters and how they are used in radar signal processing.

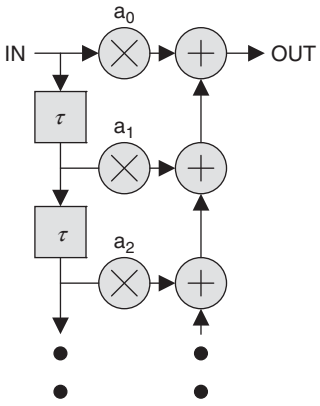


FIGURE 25.30 General direct-form FIR filter block diagram

Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) Filters. Figure 25.30 shows a block diagram of a direct-form digital FIR filter. The input sample feeds a shift register, where each block labeled τ indicates a one-sample delay in the shift register. The input sample and the output of each stage of the shift register are multiplied by unique coefficients, and the multiplier outputs are summed to produce the filtered output. Software tools exist that generate these coefficients and the number required when the user provides the desired filter characteristics, such as filter type (lowpass, highpass, bandpass, etc.), sample rate, cutoff and stopband frequency, desired passband ripple, and stopband attenuation. The filter shown is referred to as a real FIR filter, since the input data and coefficients are real values and real mathematical operations are performed. In a complex FIR filter, the data samples, coefficients, and math are complex.

This type of filter is termed *finite* impulse response because an impulse presented at the input (a single sample of "1" surrounded by samples of zeroes) would produce a finite-length output, consisting of the coefficients of the filter output in order as the "1" propagates down the shift register, as shown in Figure 25.31 for a FIR filter with seven coefficients (commonly referred to as a 7-tap FIR filter). In this example, zero-valued samples are first clocked into the FIR filter shift register, filling the shift register with zeroes and forcing the filter output to be zero. When the sample with a value of "1" is clocked into the filter, the filter output produces the first coefficient, a_0, since the other samples in the filter are still zero. On the next clock, the "1" moves to the second tap of the shift register, and a "0" is clocked into the first tap, forcing the filter output to produce the second filter coefficient, a_1. On successive clocks, the "1" propagates through the shift register, while zeroes are clocked into the shift register input, producing all of the filter coefficients on the output in sequence. The FIR filter uses feed-forward terms only, meaning that the output values only depend on the input values with no feedback terms.

Figure 25.32 depicts the general form for an infinite impulse response (IIR) filter. IIR filters make use of feed-forward and feedback terms. They are referred to as *infinite* impulse response because an impulse presented at the input to the filter will produce an infinite string of nonzero outputs in an ideal situation.

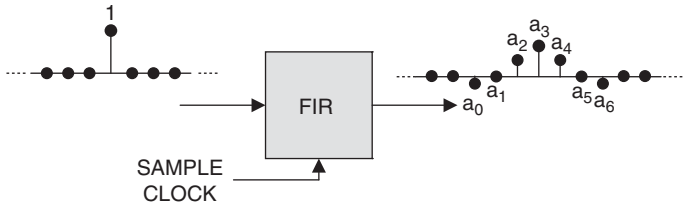


FIGURE 25.31 Impulse response of 7-tap FIR filter

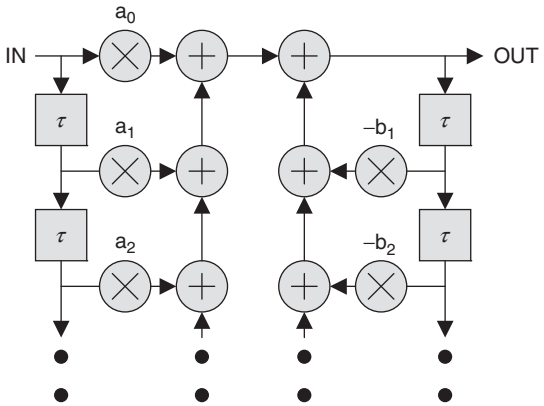


FIGURE 25.32 General IIR filter block diagram

Compared to FIR filters, IIR filters offer several advantages. In general, they require less processing and memory to implement similar functions. It is also easier to implement some filter responses as IIR rather than FIR filters. However, without careful design, IIR filter responses can be very sensitive to coefficient quantization limitations and could exhibit a tendency to *overflow* (i.e., produce an output that exceeds the processor dynamic range, determined by the number of bits in the datapath). Although IIR filters are almost never used in radar systems for these and a variety of historical reasons, a cautious designer might find an application where they can be used to good advantage.

By contrast, FIR filters are inherently stable. Real FIR filters with symmetric coefficients automatically provide a linear phase shift over frequency, introducing little or no phase distortion to the filtered signal, which is highly desirable in many applications. Because FIR filters require no feedback, they are easier to use in very high-speed applications than IIR filters, which typically require the computation of an output sample before the next output sample can be formed. Complex FIR filters, where a complex multiplication is performed at each tap, can be used to implement equalization filters, time delays, and pulse compression filters.

Figure 25.33 shows an alternative form for a FIR filter, called a *transposed form* FIR filter. In this configuration, each input sample is multiplied by all of the coefficients at once, with the sample delays between the summer outputs.

If the coefficients of a FIR filter are symmetric, so that the coefficients on either side of the center of the filter are mirror images of each other (as is the case with linear phase filters), multipliers can be saved by adding the samples that get multiplied by the same coefficient first, thereby requiring about half as many multipliers, as shown in Figure 25.34 for a 7-tap example.

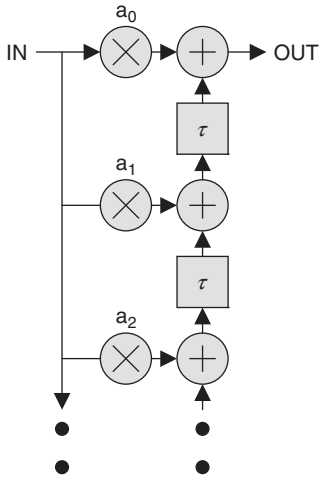


FIGURE 25.33 Transposed form FIR filter

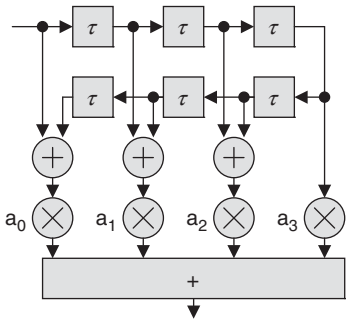


FIGURE 25.34 7-tap FIR filter with symmetric coefficients

Decimation Filters. As mentioned previously, the complexity and cost of a signal processor, in terms of the amount of system resources required to implement it, generally varies linearly with the data sample rate. For this reason, in most system applications, it is cost-effective to reduce the data sample rate to a value that is just adequate to support the bandwidth of the system. In applications where the sample rate of a signal is to be decreased (decimated), the frequency content of the signal must first be reduced so that the Nyquist criterion is satisfied for the new sample rate. This can be accomplished by first passing the signal through a digital FIR filter to restrict the bandwidth of the signal to less than half of the decimated sample rate, and then reducing the sample rate of the filtered signal by a factor of R by selecting every R^{th} sample, as described in the previous discussion of decimation. A designer can take advantage of decimation by realizing that only the filter outputs that are used need to be computed. For example, if the output of a FIR filter is to be decimated by a factor of 4, only every fourth filter output needs to be computed, which reduces the required processing by a factor of 4.

Interpolation Filters. Interpolation is the process by which the sample rate of a signal is increased, for example in preparing the signal to be upconverted to an IF, as shown in Figure 25.25. Interpolators are typically FIR filters with a lowpass filter response. To increase the sample rate by a factor R, R –1 zeroes are first inserted between the low-rate data samples, creating a data stream at a sample rate R times faster than the input rate (upsampling). This data stream is then passed through the lowpass FIR filter to produce the interpolated high-sample-rate output. Of course, the FIR filter must be clocked at the higher data rate. This process is illustrated in Figure 25.35 for a four times increase in sample rate.

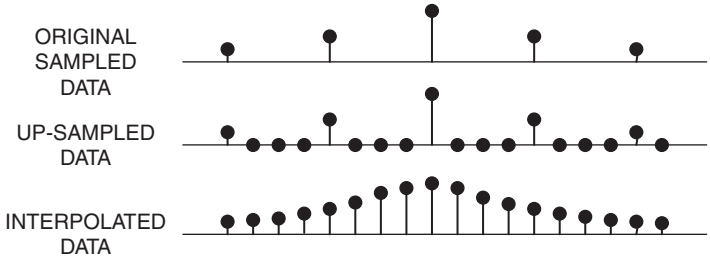


FIGURE 25.35 Illustration of interpolation filtering

Cascaded Integrator-Comb (CIC) Filters. In decimation or interpolation applications where the rate change factor is large (typically 8 or greater), a FIR filter implementation might be prohibitively costly due to the large number of filter taps that would be required. CIC filters are a class of filters introduced by Hogenauer[9] that provide a very efficient means of implementing these filter functions that do not require multipliers. Excellent descriptions of this class of filter are provided by Lyons[2] and Harris,[10] which form the basis for the following discussion.

Figure 25.36*a* shows a single-stage CIC decimator. The filter contains an integrator stage consisting of a single sample delay and an adder, followed by a comb stage with a D-stage shift register (denoted by the $D\tau$ block) and a subtractor. The comb filter gets its name because its frequency response looks like a rectified sine wave and resembles the teeth of a comb. After the comb stage, the signal is decimated by a factor R (denoted by the $\downarrow R$ block) by only passing every R^{th} sample. In most applications, the number of stages in the shift register, D, is equal to the rate change factor, R. Figure 25.36*b* depicts a CIC interpolator, where upsampling by a factor of R (denoted by the $\uparrow R$ block) is followed by a comb section and an integrator. The upsampling is accomplished by zero insertion as described in the previous section, "Interpolation Filters." Note that the processing only consists of delays and adds.

Figure 25.37*a* shows the $\sin(x)/x$ frequency response of a single-stage CIC decimator, where $R = D = 8$. The desired passband is the lightly shaded area centered at 0 Hz with bandwidth BW. The darker shaded areas with bandwidth BW in Figure 25.37*a* indicate signals that will alias into the baseband signal after decimation by 8, as shown in Figure 25.37*b* (after Lyons[2]). Note that unless BW is very small, a significant portion of out-of-band signals would get folded into the decimated baseband signal. The typical method used to improve this filter response is to increase the filter order by adding more stages. Figure 25.38 shows a three-stage CIC decimation filter, and its frequency response before and after decimation by 8 is shown in Figure 25.39*a* and *b*, respectively. Note that the aliased components are significantly reduced in amplitude, compared to the single-stage CIC filter response, and the main passband has more attenuation toward the edges. In typical applications, a CIC decimator is followed by a FIR lowpass filter and a final decimation by 2. That is, a decimate-by-16 filter would be composed of a decimate-by-8 CIC filter followed by a decimate-by-2 FIR filter. The FIR filter can be tailored to remove the undesired residual components before the final decimation. The FIR filter can also be configured to compensate for the droop in the passband response.

Figure 25.40 shows an equivalent form for a CIC decimation filter, where the decimation occurs right after the integrator section and before the comb section. The delay in the comb filter becomes a value $N\tau$, where N is equal to D/R. This allows the comb section to operate at the decimated sample rate, which makes it simpler to implement. Due to this simplification, CIC decimators are generally implemented in this form.

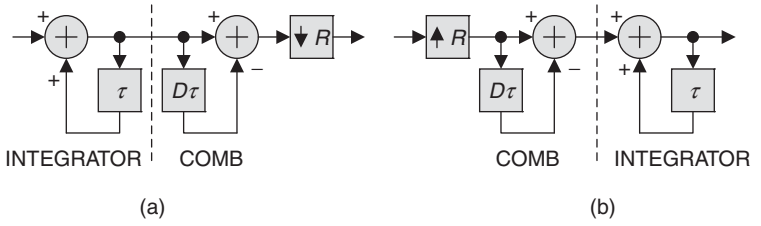


FIGURE 25.36 (*a*) CIC decimation filter and (*b*) CIC interpolation filter

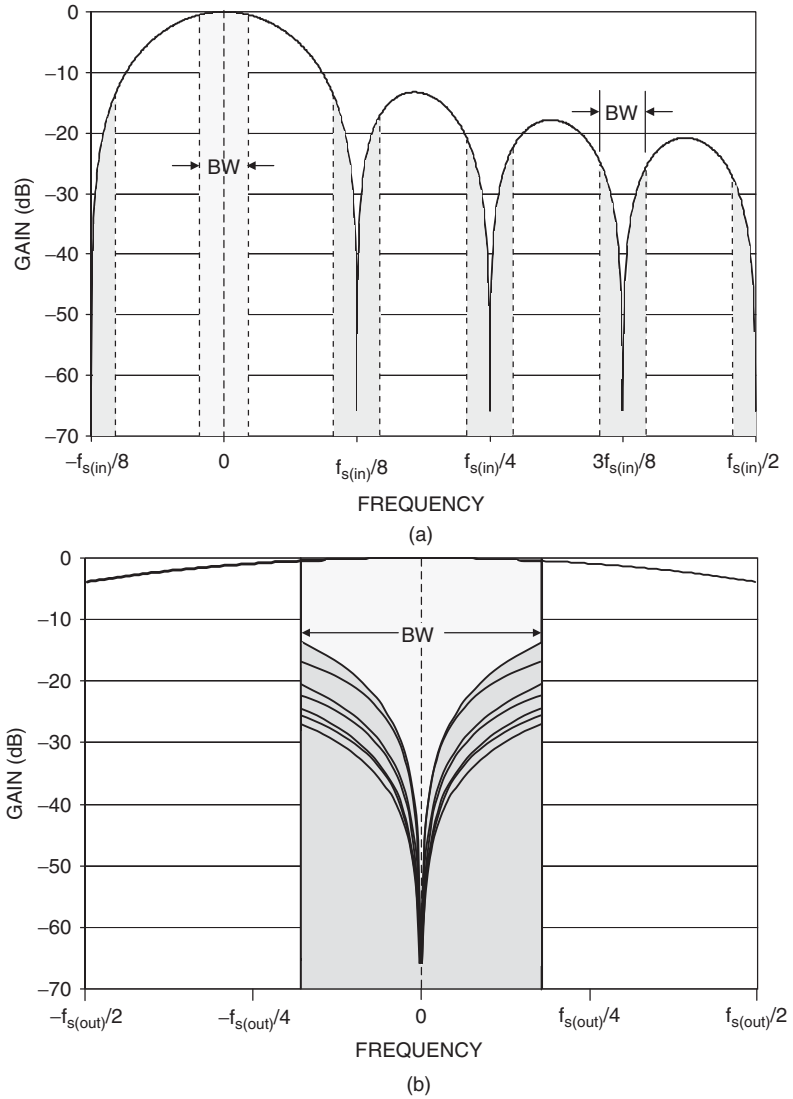


FIGURE 25.37 Frequency response of single-stage CIC decimation filter (*a*) before decimation and (*b*) after decimation

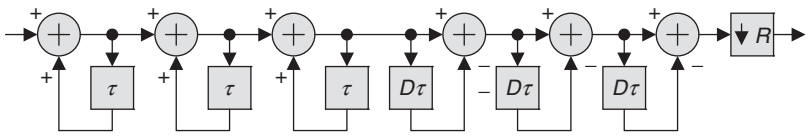


FIGURE 25.38 Three-stage CIC decimation filter

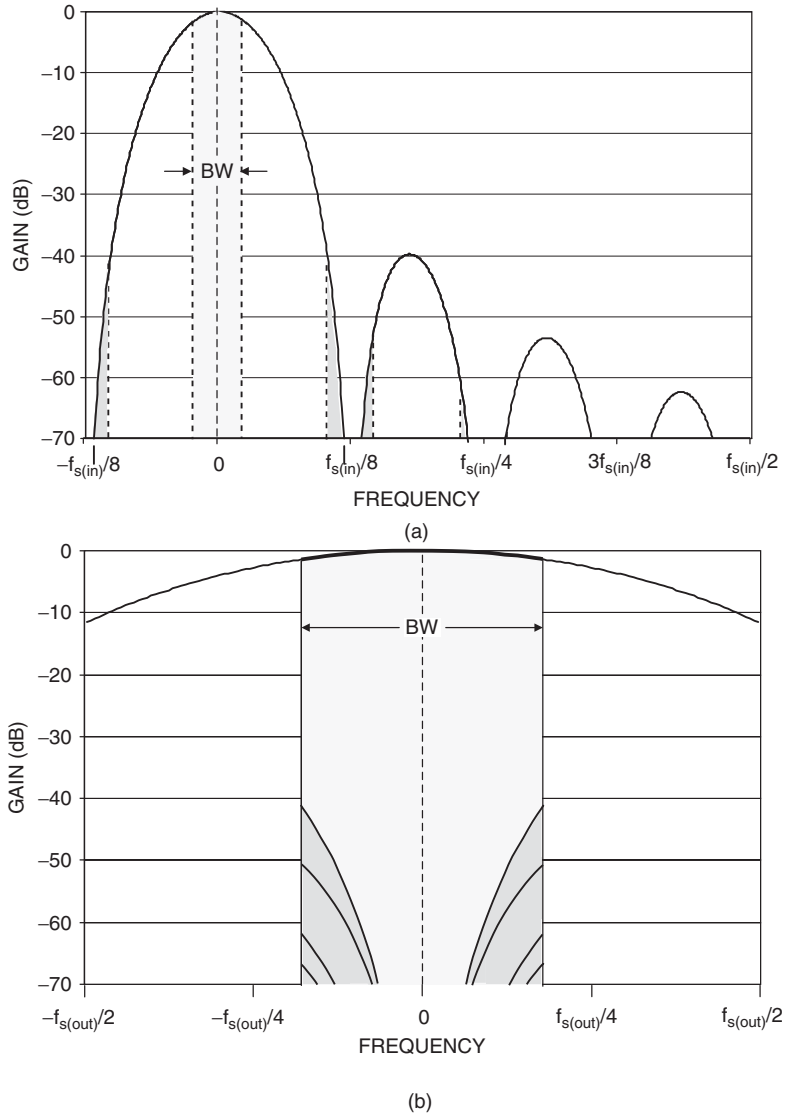


FIGURE 25.39 Frequency response of third-order CIC decimation filter (*a*) before decimation and (*b*) after decimation

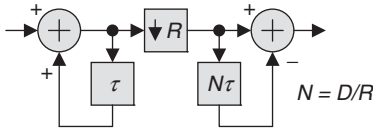


FIGURE 25.40 CIC filter with decimation after integrator

Careful inspection of the decimator architecture reveals a potential problem with the integrator. The input samples continually get added to the running sum, producing a definite overflow condition. The beauty of the architecture is that overflows are allowed and compensated for by the comb section, as long as there are enough bits in the adders to represent the maximum expected output value and the filter is implemented using two's complement arithmetic. As described by Harris,[10] the number of bits required in the adders (b_{ADDER}) is given by

$$b_{ADDER} = b_{DATA} + \text{CEIL}[\log_2(\text{GAIN})]$$

where b_{DATA} is the number of bits in the input data and CEIL[ ] indicates rounding the number in the brackets to the next highest integer. GAIN is given by

$$\text{GAIN} = R^K$$

where R is the decimation factor and K is the number of stages in the filter, resulting in

$$b_{ADDER} = b_{DATA} + \text{CEIL}[\log_2(R^K)]$$

For example, assume we have 12-bit input data ($b_{DATA} = 12$) and a 3-stage CIC filter ($K = 3$) that decimates the sample rate by a factor of 10 ($R = 10$). Substituting into this equation produces

$$b_{ADDER} = 12 + \text{CEIL}[\log_2(10^3)] = 12 + \text{CEIL}[9.966] = 12 + 10 = 22$$

In practice, although the first adder stage must support this number of bits, lower order bits may be pruned from the adders in successive stages, as described by Harris.[10]

A CIC interpolating filter would be preceded by a FIR-filter-based interpolator. CIC interpolators are described in detail in the referenced literature.

The Discrete Fourier Transform (DFT). In many sampled data systems, spectral analysis is implemented by performing the discrete Fourier transform (DFT). The DFT forms the basis for many radar signal processing algorithms, such as doppler processing and fast convolution pulse compression (described in Chapter 8), as well as radar functions such as synthetic aperture radar (SAR) and inverse synthetic aperture radar (ISAR). The DFT takes N data samples (real or complex) as input and generates N complex numbers as output, where the output samples represent the frequency content of the input data sequence. For a sample rate f_s, each output frequency sample (bin) has a width of f_s/N. The m^{th} output sample, $X(m)$, represents the amplitude and phase of the frequency content of the finite-length input sequence centered at the frequency mf_s/N.

If an input signal is exactly centered in one of the DFT frequency bins, the output will have a maximum value for that bin and nulls for all other bins. However, any frequency other than one centered in a bin will bleed into the other bins. The basic DFT bin has a frequency response similar to $\sin(x)/x$, which means that a signal in another bin might bleed into a DFT frequency bin with an attenuation as small as 13 dB. To compensate for this, the input samples can be weighted in amplitude with a wide selection of weights, such as Hanning and Hamming weights, which broaden the main lobe response of the DFT output, but reduce the amplitude of the side lobes. A thorough treatment of DFT weighting functions and their effects is given by Harris.[11]

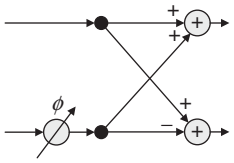


FIGURE 25.41 Radix-2 butterfly

The Fast Fourier Transform (FFT). The implementation of a DFT is computationally intensive, requiring N^2 complex multiplies. The fast Fourier transform (FFT)[12] is a very efficient technique to implement the DFT, if N is a power of 2, which requires only $(N/2)\log_2 N$ complex multiplies.

The basic computational element in an FFT is the *butterfly,* shown in Figure 25.41. In the butterfly operation, one input is phase shifted and then added to and subtracted from a second input to form two outputs. This structure is referred to as a radix-2 butterfly because it has two inputs. For certain FFT configurations, radix-4 and higher-radix butterflies provide some computational savings.

Figure 25.42 shows a radix-2, 8-point FFT. The phase shifts are represented as complex weights W_N^k, where N is the number of points in the FFT and k indicates the particular phase shift applied. W_N^k denotes a phase shift of $2k\pi/N$. These weights are often referred to as *twiddle factors*. Figure 25.43 shows the phase shifts associated with various twiddle factors.[2]

Note that the 8-point FFT consists of three stages. All of the computations in each stage are executed before proceeding to the next stage. Also note that the phase shift in the first stage, W_8^0, is zero, which requires no computation at all.

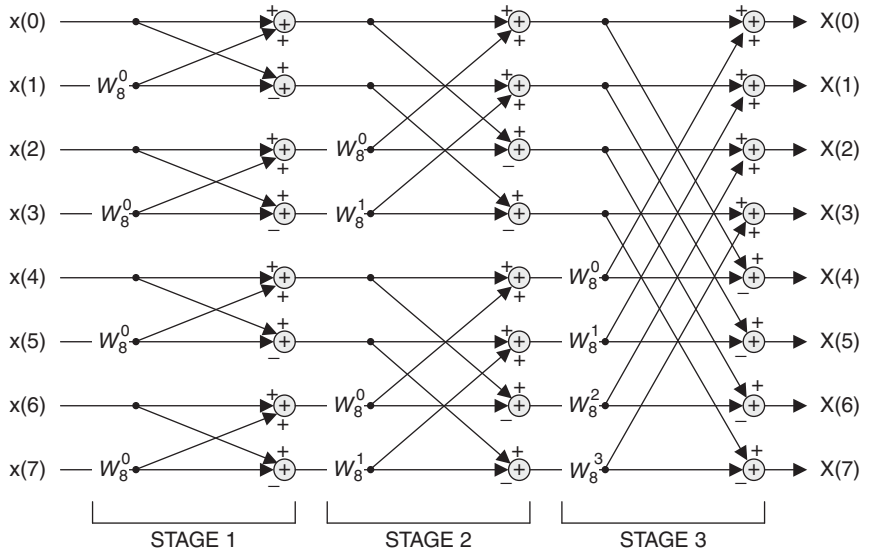


FIGURE 25.42 Eight-point, Radix-2 FFT

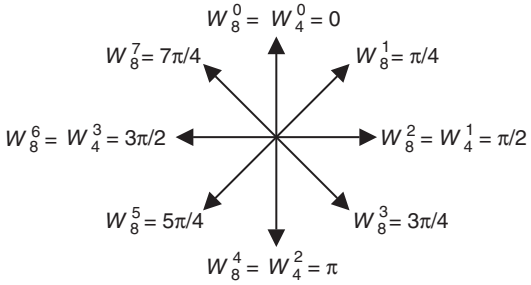


FIGURE 25.43 Phase shifts inferred by various twiddle factors

Since additions are performed in each stage, the magnitude of each output stage sample could be a factor of two or more greater than the input samples. If fixed-point computations are used, then this increased dynamic range results in a growth in the number of bits required to represent the values, and there needs to be a strategy to accommodate it.

There are several techniques generally used to handle this increased dynamic range in fixed-point FFTs. One scheme would be to ensure that the computation stages carry enough bits to accommodate the bit-growth. For example, in our 8-point FFT example, if we assume that the input samples are 12-bit complex numbers, and if we assume that the magnitudes of the complex numbers do not exceed 12 bits, then the final FFT outputs could grow 3 bits compared to the inputs, so the FFT computations could be performed with 15 bit or larger adders. This also means that the multipliers would have to handle the larger number of bits on the inputs. This method could get unwieldy for large FFTs.

Another technique is to automatically scale the outputs of each stage by a factor of 0.5, which would not allow the outputs to grow. Unfortunately, this would also limit any processing gain that the FFT might offer.

A third method, called *block floating point,* checks the magnitudes of all the outputs after each stage is computed and provides a single exponent for all output values. If any of the outputs have overflowed or come near to overflowing, then all of the outputs are scaled by a factor of 0.5, and the common exponent is incremented by 1. Enough bits have to be provided in the final mantissa to accommodate the dynamic range growth. This technique is popular because it only scales the output values when absolutely necessary.

25.5 DESIGN CONSIDERATIONS

This section addresses topics that need to be considered in the design of radar DSP systems as well as implementation alternatives.

Timing Dependencies. In coherent radar systems, all local oscillators (LOs) and clocks that generate system timing are derived from a single reference oscillator. However, this fact alone does not ensure that the transmitted waveform starts at the same RF phase on every pulse, which is a requirement for coherent systems.

Consider a system with a 5-MHz reference oscillator, from which is derived a 75 MHz IF center frequency (on transmit and receive) and a complex sample rate of 30 MHz. A rule of thumb is that the clock used to produce the pulse repetition interval (PRI) needs to be a common denominator of the IF center frequencies on transmit and receive and the complex sample frequency in order to assure pulse-to-pulse phase coherency. For this example, with an IF center frequency of 75 MHz and a complex sample rate of 30 MHz, allowable PRI clock frequencies would include 15 MHz and 5 MHz.

Hardware Implementation Technology. In the past, implementing a real-time radar digital signal processor typically required the design of a custom computing machine, using thousands of high performance integrated circuits (ICs). These machines were very difficult to design, develop, and modify. Digital technology has advanced to the point where several implementation alternatives exist that make the processor more programmable and, hence, easier to design and change.

Parallel General-purpose Computers. This architecture employs multiple general-purpose processors that are connected via high-speed communication networks. Included in this class are high-end servers and embedded processor architectures. Servers are typically homogeneous processors, where all of the processing nodes are identical and are connected by a very high-performance data bus architecture. Embedded processor architectures are typically composed of single-board computers (blades) that contain multiple general-purpose processors and plug into a standard backplane architecture, such as VME. This configuration offers the flexibility of supporting a heterogeneous architecture, where a variety of different processing blades or interface boards can be plugged into the standard backplane to configure a total system. At this writing, backplanes are migrating from parallel architectures, where data is typically passed as 32- or 64-bit words, to serial data links, which pass single bits at very high clock rates (currently in excess of 3 gigabits per second (Gbps)). These serial data links are typically point-to-point connections. In order to communicate with multiple boards, the serial links from each board go to a high-speed switch board that connects the appropriate source and destination serial links together to form a *serial fabric.* Examples of popular serial fabric backplanes at this writing include VXS, VPX, and ATCA. It is apparent that high-speed serial links will be the primary communication mechanism for multiprocessor machines into the future, with ever-increasing data bandwidths.

These parallel processor architectures offer the benefit of being programmable using high-level languages, such as C and C++. A related advantage is that programmers can design the system without knowing the intimate details of the hardware. Also, the software developed to implement the system can typically be moved relatively easily to a new hardware architecture as part of a technology refresh cycle.

On the negative side, these systems can be difficult to program to support real-time signal processing. The required operations need to be split up appropriately among the available processors, and the results need to be properly merged to form the final result. A major challenge in these applications is to support the processing *latency* requirements of the system, which defines the maximum length of time allowed to produce a result. The latency of a processor is defined as the amount of time required to observe the effect of a change at a processor's input on its output. Achieving latency goals often requires assigning smaller pieces of the workload to individual processors, leading to more processors and a more expensive system. Another challenge facing

these systems in a radar application is reset time. In a military application, when a system needs to be reset in order to fix a problem, the system needs to come back to full operation in a very short period of time. These multiprocessor systems typically take a long time to reboot from a central program store and, hence, have difficulty meeting reset requirements. Developing techniques to address these deficiencies is an active area of research. Finally, these processors are generally used for non-real-time or near-real-time data processing, as in target tracking and display processing. Since the 1990s, they have started to be applied to real-time signal processing applications. Although they might be cost-effective for relatively narrowband systems, their use in wideband DSP systems in the early 21st century is typically prohibitively expensive due to the large number of processors required. This situation should improve over time as faster and faster processors become available.

Custom-designed Hardware. Through the 1990s, real-time radar DSP systems were built using discrete logic. These systems were very difficult to develop and modify, but in order to achieve the required system performance, it was the only option available. Many systems were built using Application-Specific Integrated Circuits (ASICs), which are custom devices designed to perform a particular function. The use of ASICs allowed DSP systems to become very small with high performance. However, they were (and still are) difficult and expensive to develop, often requiring several design iterations before the device was fully operational. If an ASIC-based system needs to be modified, the ASICs need to be redesigned, incurring significant expense. Typically, the use of ASICs makes sense if tens or hundreds of thousands of units are to be sold, so that the development costs can be amortized over the life of the unit. This is rarely the case for radar systems. However, many ASICs have been developed to support the communication industry, such as digital up- and downconverters, which can be utilized in radar systems.

The introduction of the Field Programmable Gate Array (FPGA) in the 1980s heralded a revolution in the way real-time DSP systems were designed. FPGAs are integrated circuits that consist of a large array of configurable logic elements that are connected by a programmable interconnect structure. At the time of this writing, FPGAs can also incorporate hundreds of multipliers that can be clocked at rates up to a half billion operations per second, and memory blocks, microprocessors, and serial communication links that can support multigigabit-per-second data transfers. Circuits are typically designed using a hardware description language (HDL), such as VHDL (VHSIC Hardware Description Language) or Verilog. Software tools convert this high-level description of the processor to a file that is sent to the device to tell it how to configure itself. High-performance FPGAs store their configuration in volatile memory, which loses its contents when powered down, making the devices infinitely reprogrammable.

FPGAs allow the designer to fabricate complex signal processing architectures very efficiently. In typical large applications, FPGA-based processors can be a factor of ten (or more) smaller and less costly than systems based on general-purpose processors. This is due to the fact that most microprocessors only have one or very few processing elements, whereas FPGAs have an enormous number of programmable logic elements and multipliers. For example, to implement a 16-tap FIR filter in a microprocessor with a single multiplier and accumulator, it would take 16 clock cycles to perform the multiplications. In an FPGA, we could assign 16 multipliers and 16 accumulators to the task, and the filter could be performed in one clock cycle.

In order to use an FPGA most efficiently, we have to take advantage of all of the resources it offers. These include not only the large numbers of logic elements, multipliers, and memory blocks, but also the rate at which the components can be clocked. In the previous example, assume that the data sample rate is 1 MHz and also assume that the multipliers and logic can be clocked at 500 MHz. If we simply assign one multiplier to each coefficient, we would use 16 multipliers clocking at 500 MHz. Since the data rate is only 1 MHz, each multiplier would only perform one significant multiplication every microsecond and then be idle for the other 499 clocks in the microsecond, which is very inefficient. It would be much more efficient, in this case, to use one multiplier to perform as many products as possible. This technique, called *time-domain multiplexing,* requires additional logic to control the system and provide the correct operands to the multiplier at the right time. Since an FPGA can incorporate hundreds of multipliers, one can appreciate the power of this technique.

On the negative side, utilizing an FPGA to its best advantage typically requires the designer to have a thorough understanding of the resources available in the device. This typically makes efficient FPGA-based systems harder to design than systems based on general-purpose processors, where a detailed understanding of the processor architecture is not necessarily required. Also, FPGA designs tend to be aimed at a particular family of devices and take full advantage of the resources provided by that family. Hardware vendors are constantly introducing new products, invariably incorporating new and improved capabilities. Over time, the older devices become obsolete and need to be replaced during a *technology refresh* cycle. When a technology refresh occurs several years down the road, typically the available resources in the latest FPGAs have changed or a totally different device family is used, which probably requires a redesign. On the other hand, software developed for general-purpose processors may only need to be recompiled in order to move it to a new processor. Tools currently exist that synthesize C or Matlab code into an FPGA design, but these tools are typically not very efficient. The evolution of design tools for FPGAs to address these problems is an area of much research and development.

Hybrid Processors. Although it would be very desirable to simply write C code to implement a complex radar signal processor, the reality in the early 21st century is that, for many systems, implementing such a system would be prohibitively expensive or inflict major performance degradation. Although the steady increase in processor throughput may someday come to the rescue, the reality at this writing is that high-performance radar signal processors are usually a hybrid of application-specific and programmable processors. Dedicated processors, such as FPGAs or ASICs, are typically used in the high-speed front end of radar signal processors, performing demanding functions such as digital downconversion and pulse compression, followed by programmable processors in the rear, performing the lower-speed tasks such as detection processing. The location of the line that separates the two domains is application-dependent, but over time, it is constantly moving toward the front end of the system.

25.6 SUMMARY

The purpose of this chapter was to provide an overview of how digital signal processing has transformed radar system design and to give some insight into the techniques and tradeoffs that a designer has to consider. With manufacturers continually producing

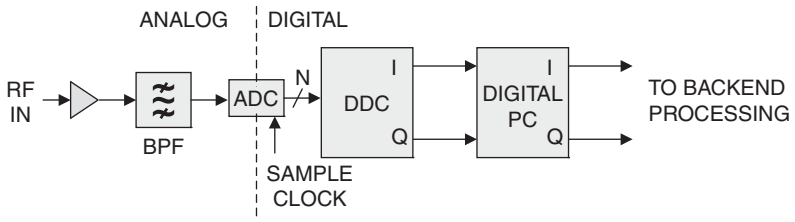


FIGURE 25.44 Direct-sampling radar digital receiver

faster and more powerful ADCs, DSP devices, and general-purpose processors, more and more of the radar system front end will move from analog to digital designs. For example, Figure 25.2 shows a typical digital receiver for a radar front end, which requires two stages of analog downconversion to bring the RF signal down to an IF that can be sampled by an ADC. This is required because of the characteristics of the ADC, which typically has poorer signal-to-noise ratio (SNR) and spur-free dynamic range (SFDR) when the input analog signal is too high, as would be the case if it were presented with the RF or high-IF signal directly. However, when faster ADCs become available, which can accommodate higher analog input frequencies while providing adequate SNR and SFDR, systems will be designed that sample the RF directly, as shown in Figure 25.44. At this writing, ADC technology allows direct sampling systems with respectable performance to be designed for radars in the HF and VHF bands. Doubtless, future components will extend this performance to higher RF frequencies.

ACKNOWLEDGMENTS

The authors would like to acknowledge the efforts of and extend their sincere gratitude to several individuals who helped them immensely in the preparation of this chapter. First, to Mr. Gregory Tavik of NRL for his thorough review of this chapter and the many excellent comments he made. Next, to Dr. Fred Harris of San Diego State University and Mr. Richard Lyons, who graciously reviewed sections of the chapter and offered several suggestions, all of which were incorporated.

REFERENCES

1. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, 2nd Ed., Englewood Cliffs, NJ: Prentice-Hall, 1989.
2. R. G. Lyons, *Understanding Digital Signal Processing*, 2nd Ed., Upper Saddle River, NJ: Prentice Hall, 2004.
3. J. O. Coleman, "Multi-rate DSP before discrete-time signals and systems," presented at First IEEE Workshop on Signal Processing Education (SPE 2000), Hunt, TX, October 2000.
4. W. M. Waters and B. R. Jarrett, "Bandpass signal sampling and coherent detection," *IEEE Trans. On Aerospace Electronic Systems,* vol. AES-18, no. 4, pp. 731–736, November 1982.
5. D. P. Scholnik and J. O. Coleman, "Integrated I-Q demodulation, matched filtering, and symbol-rate sampling using minimum-rate IF sampling," in *Proc. of the 1997 Symposium on Wireless Personal Communication*, Blacksburg, VA, June 1997.

6. B. Brannon and A. Barlow, "Aperture uncertainty and ADC system performance," Analog Devices Application Note AN-501, Rev. A, March 2006.
7. J. E. Volder, "The CORDIC trigonometric computing technique," *IRE Trans. on Electronic Computers*, vol. EC-8, pp. 330–334, 1959.
8. R. Andraka, "A survey of CORDIC algorithms for FPGA-based computers," in *ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, Monterey, CA, February 1998, pp. 191–200.
9. E. B. Hogenauer, "An economical class of digital filters for decimation and interpolation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-29(2), pp. 155–162, April 1981.
10. F. J. Harris, *Multirate Signal Processing for Communication Systems,* Upper Saddle River, NJ: Prentice Hall, 2004.
11. F. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, January 1978, pp. 51–83.
12. J. Cooley and J. Tukey, "An Algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, April 1965.

Chapter 26

The Propagation Factor, F_p, in the Radar Equation

Wayne L. Patterson

Space and Naval Warfare Systems Center
Atmospheric Propagation Branch

26.1 INTRODUCTION

Paralleling the development of radar technology is the development of the radar equation. As engineering considerations such as the probability of detection, probability of false alarm, signal loss factors, and signal-to-noise ratio allowed the radar range equation to develop sufficiently to be useful in radar performance analysis, developing computer technologies allowed for more sophisticated radar range equation solution techniques. Thus, solution techniques to the radar range equation migrated from pencil and paper "worksheets" to simple computer programs automating the "worksheet" to highly sophisticated computer programs accounting for technology advances in signal processing and environmental modeling.

The intent of this chapter is twofold. The first is to focus on one particular term of the radar range equation, the propagation factor F_P (defined in Section 26.6). Encompassed within the propagation factor are all the effects upon propagation attributable to the natural environment. These effects are energy absorption from gasses and liquid water, diffraction, refraction, multipath interference, earth-surface dielectrics, terrain interference, and a number of other natural environmental considerations.

The second focus of this chapter is to describe the computer modeling of the propagation factor. For ease of computations in early solution techniques, the propagation factor was often taken as unity, a condition representing free space. With computer-implemented propagation models, however, the assumption of free space need no longer be a limiting factor. One such propagation model, the Advanced Propagation Model (APM) and its graphical user interface program, the Advanced Refractive Effects Prediction System (AREPS),[1] are featured here. While the focus upon AREPS within this chapter is for the understanding of how important the propagation factor is within the radar equation, AREPS is much more than a propagation factor tool. AREPS provides the radar engineer and the operational radar operator with an easy to use but extremely powerful method to define the natural atmospheric environment using data from a wide range of sources; to manage, create, and define various elements of terrain data; to execute the appropriate propagation model for the task at hand; and then to present the results in a number of different and highly configurable graphic and

text displays, including exporting computed data in several formats for import into other applications. AREPS is not limited to just radar applications, however. AREPS together with APM and its other embedded propagation models can provide assessments for LF to EHF communications (ground and sky wave), strike and electronic countermeasures, Electronic Support Measures (ESM) vulnerabilities, and many other applications. AREPS and APM are products of the atmospheric propagation branch of the Space and Naval Warfare Systems Center (SPAWARSYSCEN), San Diego. AREPS will execute on a personal computer (desktop or laptop) using a Microsoft Windows operating system such as NT, 2000, XP, or Vista and requires no additional special hardware. AREPS may be freely obtained at the URL listed in the first reference.

Before continuing with the discussion of electromagnetic (EM) propagation models and assessment systems, it is appropriate to discuss the natural environment and its influence upon EM system performance.

26.2 THE EARTH'S ATMOSPHERE[1]

Structure and Characteristics. The Earth's atmosphere is a collection of many gases together with suspended particles of liquids and solids. Excluding variable components such as water vapor, ozone, sulfur dioxide, and dust, the gases of nitrogen and oxygen occupy about 99% of the volume, with argon and carbon dioxide being the next two most abundant gases. From the Earth's surface to an altitude of approximately 80 kilometers, mechanical mixing of the atmosphere by heat-driven air currents evenly distributes the components of the atmosphere. At about 80 kilometers, the mixing decreases to the point where the gases tend to stratify in accordance with their weights.

The lower, well-mixed portion of the atmosphere is called the *homosphere*, while the higher, stratified portion is called the *heterosphere*. Within the heterosphere lies the *ionosphere*. The bottom portion of the homosphere is called the *troposphere*.

Troposphere. The troposphere extends from the Earth's surface to an altitude of 8 to 10 kilometers at polar latitudes, 10 to 12 kilometers at middle latitudes, and up to 18 kilometers at the equator. It is characterized by a temperature decrease with height. The point at which the temperature ceases to decrease with height is known as the *tropopause*. The average vertical temperature gradient of the troposphere varies between 6° and 7° Celsius per kilometer.

The concentrations of gas components of the troposphere vary little with height, except for water vapor. The water vapor content of the troposphere comes from evaporation of water from oceans, lakes, rivers, and other water reservoirs. Differential heating of land and ocean surfaces produces vertical and horizontal wind circulations that distribute the water vapor throughout the troposphere. The water vapor content of the troposphere rapidly decreases with height. At an altitude of 1.5 kilometers, the water vapor content is approximately half of the surface content. At the tropopause, the content is only a few thousandths of what it is at the surface.

In 1922, the Weather Bureau, at the request of the National Advisory Committee for Aeronautics (NACA), prepared a standard atmosphere for scientific and engineering use based primarily on the average conditions over the United States at latitude 40°. In 1925, the computations were extended to 20,000 meters using constants adopted by the NACA. An extension of the standard atmosphere to 120,000 meters was prepared in 1947.

The standard atmosphere is based primarily on the assumption of a linear decrease in temperature with height up to the tropopause and an isothermal layer above. In addition, certain other assumptions are

(a) The air is dry.

(b) The air is a perfect gas, obeying the Laws of Charles and Boyle.

(c) Gravity is constant at all altitudes.

(d) The temperature of the isothermal atmosphere is −55° C.

(e) The linear decrease of temperature with height is −6.5° C per kilometer.

The International Commission for Air Navigation (ICAN) uses the 1924 NACA standard atmosphere, with minor modifications, primarily in the value of gravity and the temperature of the isothermal region. For generic radar studies and other radar applications, such as target height calculations for height-finding radars, it is the propagation through this standard atmosphere that is considered.

26.3 REFRACTION[2]

Index of Refraction. The term *refraction* refers to the property of a medium to bend an electromagnetic wave as it passes through the medium. A measure of the amount of refraction is the index of refraction, n, defined as the velocity, c, of propagation in free-space (away from the influence of the Earth or other objects) to the velocity, v, in the medium. This is

$$n = \frac{c}{v} \tag{26.1}$$

Refractivity and Modified Refractivity in the Troposphere. The normal value of the refractive index, n, for the atmosphere near the Earth's surface varies between 1.000250 and 1.000400. For studies of propagation, the index of refraction is not a very convenient number; therefore, a scaled index of refraction, N, called *refractivity*, has been defined. At microwave frequencies and below, the relationship between the index of refraction, n, and refractivity, N, for air that contains water vapor is given as

$$N = (n-1)10^6 = \frac{77.6p}{T} + \frac{e_s\, 3.73 \times 10^5}{T^2} \tag{26.2}$$

where e_s is the partial pressure of water vapor in millibars or

$$e_s = \frac{rh\ 6.105\ e^x}{100} \tag{26.3}$$

$$x = 25.22\frac{T-273.2}{T} - 5.31\ \log_e\left(\frac{T}{273.2}\right) \tag{26.4}$$

p = atmosphere's barometric pressure in millibars
T = atmosphere's absolute temperature in Kelvin
rh = atmosphere's relative humidity in percent

Thus, the atmospheric refractivity near the Earth's surface would normally vary between 250 and 400 *N*-units.

Because the barometric pressure and water vapor content of the atmosphere decrease rapidly with height while the temperature decreases slowly with height, the index of refraction, and therefore refractivity, normally decreases with increasing altitude.

While a radar engineer may like to consider refraction in terms of *N*-units because it provides a better physical point of view, an AREPS user may not be a radar engineer but a tactical operator such as a combat pilot. In graphically examining refractive gradients and their effect upon propagation (such as ducting described in Section 26.5), a modified refractivity, defined as

$$M = N + 0.157\,h \quad \text{for altitude } h \text{ in meters an} \tag{26.5}$$

$$M = N + 0.048\,h \quad \text{for altitude } h \text{ in feet} \tag{26.6}$$

is used in place of the refractivity. While a graphical *N*-unit versus height display will show a negative slope (decreasing *N*-units) with height, a graphical *M*-unit versus height display will show a change in slope, from positive (increasing *M*-units) under standard atmospheric conditions to a negative slope (decreasing *M*-units) under ducting atmospheric conditions. Therefore, a *M*-unit type display is more readily understood by the tactical radar operator looking for an optimum flight altitude for his attack.

26.4 STANDARD PROPAGATION[2]

Standard propagation mechanisms are those mechanisms and processes that occur in the presence of a standard atmosphere. These propagation mechanisms are standard refraction, free-space propagation, multipath interference (or surface reflection), diffraction, and tropospheric scatter.

Normal/Standard Refraction. The refractivity distribution within the atmosphere is nearly an exponential function of height. The decrease of *N* with height close to the Earth's surface (within 1 kilometer) is sufficiently smooth, however, to allow an approximation of the exponential function by a linear function. This linear function is known as a *standard* gradient and is characterized by a decrease of 39 *N*-units per kilometer, or an increase of 118 *M*-units per kilometer. A standard gradient will cause traveling EM waves to bend downward from a straight line. Gradients that cause effects similar to a standard gradient but vary between 0 and −79 *N*-units per kilometer or between 79 and 157 *M*-units per kilometer are known as *normal* gradients.

Free-space Propagation. The simplest case of electromagnetic wave propagation is the transmission of a wave between a transmitter and a receiver in free space. Free space is defined as a region whose properties are isotropic, homogeneous, and loss-free, i.e., away from the influences of the Earth's atmosphere and surface. In free space, the electromagnetic wavefront from an isotropic radiator spreads uniformly in all directions from the transmitter.

Multipath Interference and Surface Reflection. When an electromagnetic wave strikes a nearly smooth large surface, such as the ocean, a portion of the energy is reflected from the surface and continues propagating along a path that makes an

angle with the surface equal to that of the incident wave, as shown in Figure 26.1.

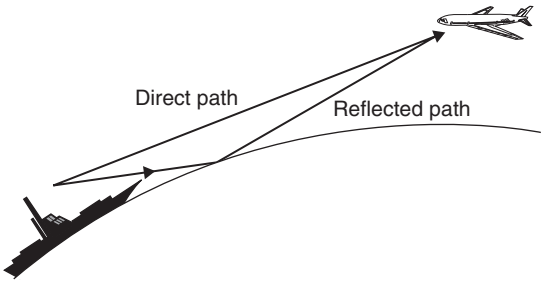


FIGURE 26.1 Surface reflection

The strength of the reflected wave is determined by the reflection coefficient, a value that depends upon the frequency and polarization of radiation, the angle of incidence, and the roughness of the reflecting surface.

For shallow incidence angles and smooth seas, typical values of the reflection coefficient are near unity (i.e., the reflected wave is almost as strong as the incidence wave). As the wind speed increases, the ocean surface grows rougher and the reflection coefficient decreases. For a transmitter near the surface, the reflection process results in two paths to a receiver within the line of sight.

As stated above, upon reflection, a portion of the energy is propagated in the direction of initial wave motion. A portion of energy is also reflected backward toward the transmitter. This backward reflected energy is also received by the radar and may interfere with the radar's ability to distinguish a desired target. This backward reflected energy is called *clutter*.

Not only is the magnitude of the reflected wave reduced, but also the phase of the wave is altered. For horizontally or vertically polarized waves at low grazing angles, the phase change upon reflection is approximately 180°. Whenever two or more wave trains traveling over different paths intersect at a point in space, they are said to interfere (multipath interference). If two waves arrive at the same point in phase, they constructively interfere, and the electric field strength is greater than either of the two component waves taken alone. If the two waves arrive together out of phase, they destructively interfere, and the resultant field strength is weakened.

As the geometry of the transmitter and receiver change, the relative lengths of the direct path and reflected path also change, which results in the direct and reflected wave arriving at the receiver in varying amounts of phase difference. The received signal strength, which is the vector sum of the signal strengths of the direct and reflected wave, may vary up to 6 dB above and 20 dB or more below the free-space value.

Diffraction. Energy tends to follow along the curved surface of an object. The degree of refraction is dependent upon the polarization of the propagating wave and the size of the diffracting object relative to the wavelength. Diffraction is the process by which the direction of propagating radiation is changed so that it spreads into the geometric shadow region of an opaque object that lies in the radiation field. In the earth-atmosphere system, diffraction occurs where the straight-line distance between the transmitter and receiver is just tangent to the Earth's surface. For a homogeneous atmosphere, this point of tangency with the Earth is referred to as the *geometrical* horizon. For an inhomogeneous atmosphere (using an effective earth radius) and at radar and optical frequencies, this point of tangency is referred to as the *radar* and *optical* horizon, respectively.

The ability of the electromagnetic wave to propagate beyond the horizon by diffraction is highly dependent upon frequency. The lower the frequency, the more the wave is diffracted. At microwave radar frequencies, the wavelength is small when compared

to the Earth's dimensions, and little energy is diffracted. At optical frequencies or very short radar wavelengths, the optical horizon represents the approximate boundary between regions of propagation and no propagation.

Tropospheric Scatter. At ranges far beyond the horizon, the propagation loss is dominated by troposcatter. Propagation in the troposcatter region is the result of scattering by small inhomogeneities within the atmosphere's refractive structure. At radar frequencies, troposcatter is generally not considered for radar range performance. However, troposcatter scattering could be an important consideration in target detection by a receiver not co-located with the radar itself or detection of the radar's emissions by an Electronic Support Measures (ESM) system.

26.5 ANOMALOUS PROPAGATION[2]

Anomalous or nonstandard wave propagation usually refers to the consideration of nonstandard refraction versus standard refraction. These nonstandard refractive conditions lead to trans-horizon paths, decreased horizon paths, and distortions in simple surface refection and multipath interference.

Subrefraction. If the motions of the atmosphere produce a situation where the temperature and humidity distribution create an increasing value of N with height, the wave path would actually bend upward and the energy would travel away from the Earth. This is termed *subrefraction*. Although this situation occurs infrequently in nature, it still must be considered when assessing electromagnetic systems' performance. For example, an Atlantic coast vessel traffic control radar located near the entrance to the Delaware Bay observed a reduction in detection range from 37 to 17 km. Sometimes ships can be seen visually from the radar tower before they can be observed on the radar screen. The reduction in the radar detection range usually lasts several hours and occurs often when fog is present.[3]

A subrefractive layer of the troposphere would cause the propagating energy to bend upward or away from the Earth's surface, thereby leading to decreased detection ranges and shortened radio horizons.

Subrefractive layers may be found at the Earth's surface or aloft. In areas where the surface temperature is greater than 30° Celsius, and relative humidities are less than 40% (i.e., large desert and steppe regions), solar heating will produce a very nearly homogeneous surface layer, often several hundreds of meters thick. Because this layer is unstable, the resultant convective processes tend to concentrate any available moisture near the top of the layer. This in turn creates a positive N gradient or subrefractive stratum aloft. This layer may retain its subrefractive nature into the early evening hours, especially if a radiation inversion develops, trapping the water vapor between two stable layers.

For areas with surface temperatures between 10° and 30° Celsius and relative humidities above 60% (i.e., the western Mediterranean, Red Sea, Indonesian Southwest Pacific), surface-based subrefractive layers may develop during the night and early morning hours. These layers are characteristically caused by advection of warm, moist air over a relatively cooler and drier surface. While the N gradient is generally more intense than that described above, the layer is often not as thick. Similar conditions may also be found in regions of warm frontal activity.

Superrefraction. If the troposphere's temperature increases with height (temperature inversion) and/or the water vapor content decreases rapidly with height, the refractivity gradient will decrease from the standard. The propagating wave will be bent downward from a straight line more than normal. As the refractivity gradient continues to decrease, the radius of curvature for the wave path will approach the radius of curvature for the Earth. The refractivity gradient for which the two radii of curvature are equal is referred to as the *critical gradient*. At the critical gradient, the wave will propagate at a fixed height above the ground and will travel parallel to the Earth's surface. Refraction between the normal and critical gradients is known as *superrefraction*.

Superrefractive conditions are largely associated with temperature and humidity variations near the Earth's surface. Inversions aloft, due to large-scale subsidence will lead to superrefractive layers aloft. Superrefractive layers will lead to an increase of radar detection ranges and extensions of the radio horizon.

The effects of a superrefractive layer upon a surface-based system are directly related to its height above the Earth's surface. For airborne systems, the effects of a superrefractive layer depend upon the position of the transmitter and receiver relative to the layer. Both of these factors are related to the electromagnetic wave's angle of layer penetration. The steeper the penetration angle, the less of an effect the layer will have upon propagation.

Trapping. Trapping is an extension of superrefraction because the meteorological conditions for both are the same. Should the refractivity gradient decrease beyond the critical gradient, the radius of curvature for the wave will become smaller than the Earth's curvature. The wave will either strike the Earth and undergo surface reflection, or enter a region of standard refraction and be refracted back upward, only to reenter the area of refractivity gradient that causes downward refraction. This refractive condition is called trapping because the wave is confined to a narrow region of the troposphere. The common term for this confinement region is a *tropospheric duct* or a *tropospheric waveguide*. It should be noted that a tropospheric waveguide is not a waveguide in the true sense of the word because there are no rigid walls that prevent the escape of energy from the guide.

The refractivity gradients and their associated refractive conditions are summarized in Table 26.1.

Atmospheric Ducts. A duct is a channel in which electromagnetic energy can propagate over great ranges. To propagate energy within a duct, the angle made by the electromagnetic system's energy with the duct must be small, usually less than 1°. Thicker ducts, in general, can support trapping for lower frequencies. The vertical distribution of refractivity for a given situation must be considered as well as the geometrical relationship of transmitter and receiver to the duct in order to assess the duct's effect at any particular frequency.

TABLE 26.1 Refractive Gradients and Conditions

Condition	*N*-Gradient	*M*-Gradient
Trapping	< –157 *N*/km or < –48 *N*/kft	< 0 *M*/km or < 0 *M*/kft
Superrefraction	–157 to –79 *N*/km or –48 to –24 *N*/kft	0 to 79 *M*/km or 0 to 24 *M*/kft
Normal	–79 to 0 *N*/km or –24 to 0 *N*/kft	79 to 157 *M*/km or 24 to 48 *M*/kft
Standard	–39 *N*/km	118 *M*/km
Subrefraction	> 0 *N*/km or > 0 *N*/kft	> 157 *M*/km or > 48 *M*/kft

A simple relationship between a duct's thickness and its ability to trap a particular frequency is given by

$$\lambda_{\max} = 2.5 \times 10^{-3} \left(\frac{\delta N}{t} - 0.57 \right)^{0.5} t^{1.5} \tag{26.7}$$

where $\lambda_{\max}$ is the maximum frequency trapped, δN is the refractive index change across the duct, and t is the duct thickness.[4]

In addition to extended radar ranges, atmospheric ducts (and other propagation effects such as multipath interference) have other significant impacts upon radar performance. These effects may be visualized with the aid of a height versus range graphic available from an assessment system such as AREPS. Such a graphic is shown in Figure 26.2. In this figure, the different shadings correspond to different propagation loss values (defined later as computed by the propagation model). The actual values are immaterial to this illustration as the important features to note are the consequences of ducting. In this figure, one can clearly see the null and lobe structure resulting from multipath interference. While a discussion of ducting conditions upon EM wave propagation is usually concerned with propagation beyond the normal horizon, ducting can have an effect within the horizon. Ducting can alter the normal lobe pattern caused by the interference of the direct ray and the surface-reflected ray. The relative phase between the direct and reflected path may be changed as well as the relative amplitudes of the two rays. The effect of the duct on the line-of-sight propagation is to reduce the angle of the lowest lobe, bringing it closer to the surface.[5]

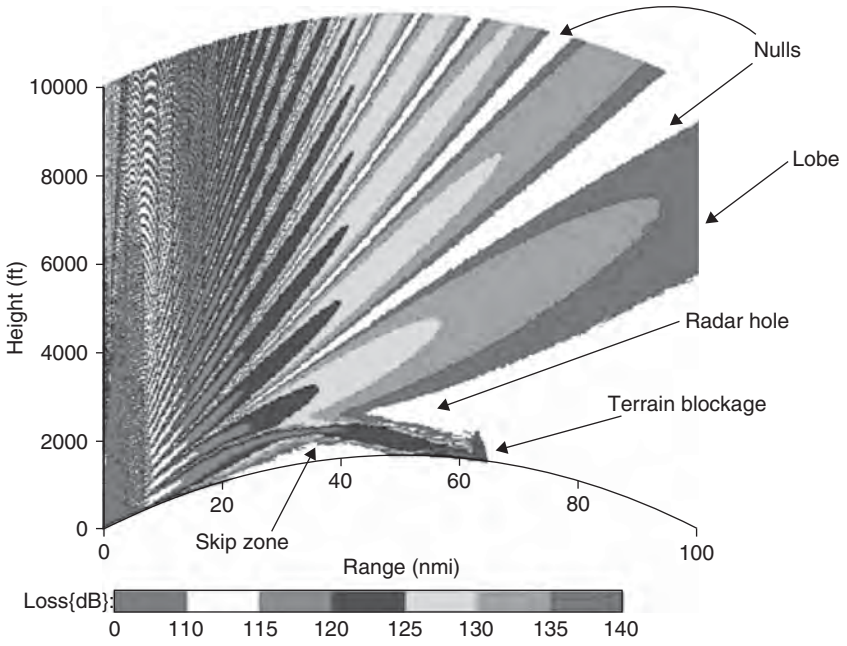


FIGURE 26.2 Ducting consequences

Ducts not only give extended radar detection ranges for systems within the duct, but they may also have a dramatic effect upon transmitter/receiver systems that transcend duct boundaries. For example, an air target that would normally be detected may be missed if the radar is within or just above the duct and the target is just above the duct. This area of reduced coverage is known as a *radar hole* or *shadow zone*.

Another interesting feature of surface-based ducts is the skip zone near the normal horizon, in which the duct has no influence. It should be noted that the surface duct created from a surface-based trapping layer does not have this skip zone phenomenon.

Height-finding radars usually determine height from energy path assumptions within a normal environment. Nonstandard refractive conditions as discussed earlier will cause the energy's path to deviate from these assumptions, resulting in errors in altitude calculations. Figure 26.3 shows the downward deviated path associated with a surface ducting condition compared to the path under normal conditions. It can be seen that actual altitude of the radar target is lower than calculated by the height-finding radar. This error could lead to significant tactical consequences in a ship's self-defense scenario.

For a practical example of ducting effects, one can read about the "Battle of the Pips."[6] During the summer of 1942, two U.S. Navy task groups were deployed to remove the Japanese occupation of Attu, Alaska. On the night of July 25th, the USS Mississippi gained radar contact with what was believed to be Japanese fleet units moving toward Attu in order to withdraw troops. Radars onboard the USS New Mexico, the USS Portland, and the USS Wichita confirmed the radar contacts. On the orders of Admiral Giffen, the U.S. Navy ships opened fire. The firing continued for about half an hour during which time 518 14-inch shells and 487 8-inch shells were expended. Gun flashes were reportedly seen by the Japanese on Kiska, 80 miles away. Two other ships, the destroyers USS San Francisco and USS Santa Fe, could not gain radar contact, but they did detect the splashes made by shells hitting the water. Nothing was ever found. When U.S. and Canadian forces landed at Attu on August 15th, they found the island abandoned. The 5,000 Japanese troops had been evacuated under cover of fog and rain on July 28th. Later investigation determined that the Japanese evacuation

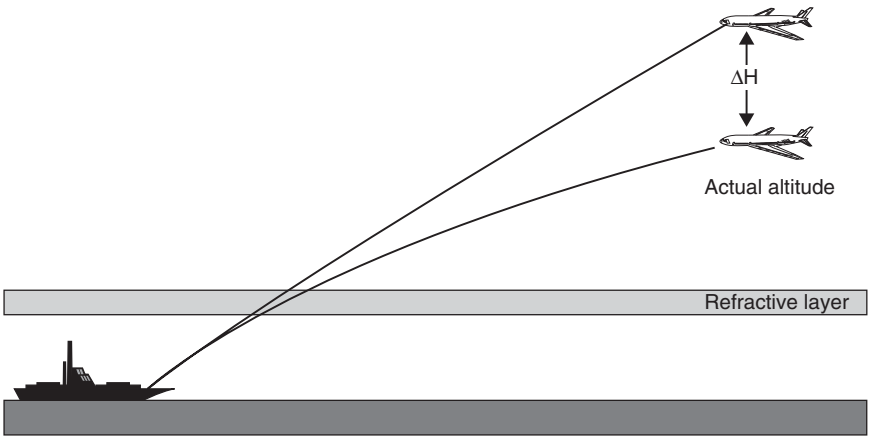


FIGURE 26.3 Altitude errors

ships were 500 nautical miles SW of Kiska during the firing. The radar contacts were actually returns from land that had been ducted over the horizon.

Several meteorological conditions will lead to the creation of ducts. Where these conditions exist and what these conditions are determines the name and nature of the duct.

Surface Ducts. If the meteorological conditions cause a trapping layer to occur, such that the base of the resultant duct is at the Earth's surface, a surface duct is formed. There are three types of surface ducts based on the trapping layer's relationship to the Earth's surface. The trapping layer is indicated graphically by the solid black *M*-unit versus height line where the slope of the line is negative (*M*-unit decrease with height).

The first type of duct is a surface duct created from a surface-based trapping layer. This duct is referred to as a *surface* duct and is illustrated in Figure 26.4. The dotted line in the figure shows the vertical dimension of the duct from bottom to top. The second type of surface duct is created from an elevated trapping layer. This duct is commonly referred to as a *surface-based* duct and is illustrated in Figure 26.5. Note the duct, identified by the dashed line, contains the trapping layer and "normal" gradient layer below. The third type of surface duct is one created by a rapid decrease of relative humidity immediately adjacent to the air-sea interface. This duct is referred to as an *evaporation* duct. Because the evaporation duct is of great importance for over-water EM propagation, it warrants a detailed discussion. This discussion appears in its own section below.

Surface-based ducts occur when the air aloft is exceptionally warm and dry compared with the air at the Earth's surface. Several meteorological conditions may lead to the formation of surface-based ducts.

Over the ocean and near land masses, warm, dry continental air may be advected over the cooler water surface. Examples of this type of advection are the Santa Ana of southern California, the Sirocco of the southern Mediterranean, and the Shamal of the Persian Gulf. This advection will lead to a temperature inversion at the surface. In addition, moisture is added to the air by evaporation, producing a moisture gradient to strengthen the trapping gradient. This type of meteorological condition routinely leads to a surface duct created by a surface-based trapping condition. However, as you travel from the coastal environment into the open ocean, this trapping layer may well rise from the surface, thereby creating the surface-based duct. Surface-based ducts tend

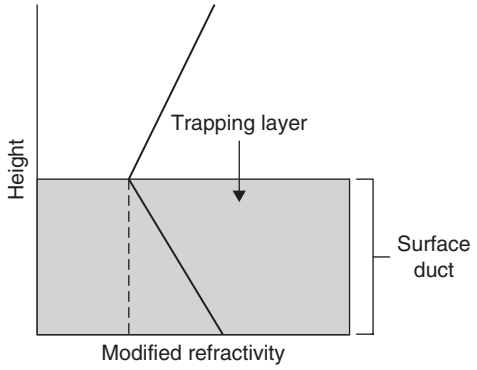


FIGURE 26.4 Surface duct

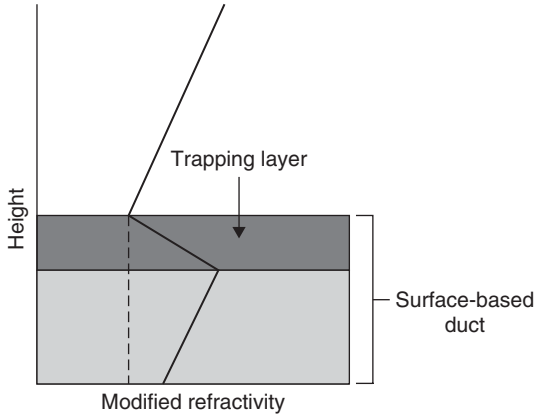


FIGURE 26.5 Surface-based duct

to be on the leeward side of land masses and may occur both during the day and at night. In addition, surface-based ducts may extend over the ocean for several hundred kilometers and may be very persistent (lasting for days).

Another method of producing surface-based ducting conditions is by divergence (spreading out) of relatively cool air under a thunderstorm. While this method may not be as frequent as the other methods, it may still enhance surface propagation during the thunderstorm activity, usually on the order of a few hours.

With the exception of thunderstorm conditions, surface-based ducting is associated with fair weather and with increased occurrence of surface-based ducts during the warmer months and in more equatorial latitudes. Any time the troposphere is well mixed, such as with frontal activity or with high wind conditions, surface-based ducting is decreased.

Evaporation Ducts. A change in the moisture distribution without an accompanying temperature change can also lead to a trapping refractivity gradient. The air in contact with the ocean's surface is saturated with water vapor. A few meters above the surface the air is not usually saturated, so there is a decrease of water vapor pressure from the surface to some value well above the surface. The rapid decrease of water vapor initially causes the modified refractivity, M, to decrease with height, but at greater heights the water vapor distribution will cause M to reach a minimum and, thereafter, increase with height. The height at which M reaches a minimum is called the *evaporation duct height*, as illustrated in Figure 26.6.

Evaporation ducts exist over the ocean, to some degree, almost all the time. The duct height varies from a meter or two in northern latitudes during winter nights to as much as 40 meters in equatorial latitudes during summer days. On a world average, the evaporation duct height is approximately 13 meters. It should be emphasized that the evaporation duct "height" is not a height below which an antenna must be located in order to have extended propagation but a value that relates to the duct's strength or its ability to trap radiation. The duct strength is also a function of wind velocity. For unstable atmospheric conditions (conditions where a cooler layer of air overlies a warmer layer of air), stronger winds generally result in stronger signal strengths (or less propagation loss) than do weaker winds.

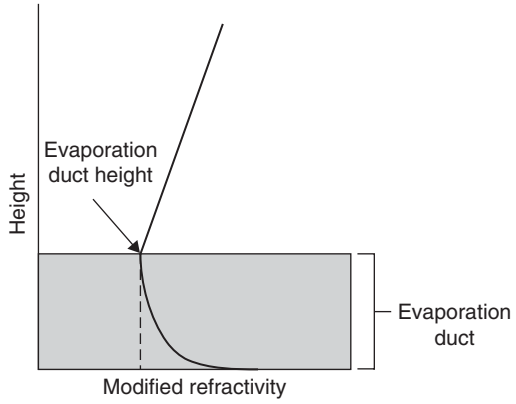


FIGURE 26.6 Evaporation duct

Because the evaporation duct is much weaker than the surface-based duct, its ability to trap energy is highly dependent on frequency. Generally, the evaporation duct is only strong enough to affect electromagnetic systems above 3000 MHz.

For surface ducting conditions, the vertical extent of the duct is sufficient to allow for its measurement using an ascending radiosonde, a descending rocketsonde, or a microwave refractometer onboard some sort of air vehicle. However, for an evaporation duct, it is not the vertical extent of the duct that is important but the refractive gradient within the duct. Changes of refractive gradients over vertical heights less than a few millimeters may have a significant impact upon the duct's trapping ability. Thus, assessment of the evaporation duct is best performed by making surface meteorological measurements and inferring the duct height from the meteorological processes occurring at the air/sea interface and not from direct measurements using the traditional radiosonde, rocketsonde, or microwave refractometer. With the advent of newer, high-resolution sondes that may be lowered to the surface from a ship, the impression is given that the evaporation duct may be measured directly. For practical applications, however, this impression is false and a direct measurement should not be attempted. Due to the turbulent nature of the troposphere at the ocean surface, a refractivity profile measured at one time would most likely not be the same as one measured at another time, even when the two measurements are seconds apart. Therefore, any measured profile would not be representative of the average evaporation ducting conditions, the conditions that an assessment system must consider.

Elevated Ducts. If meteorological conditions cause a trapping layer to occur aloft, such that the base of the duct occurs above the Earth's surface, the duct is referred to as an elevated duct, as illustrated in Figure 26.7. Note again the advantage of showing *M*-unit gradients versus *N*-unit gradients. From Figure 26.7, it can be seen that the duct extends from the top of the trapping layer downward until it intersects with the *M*-unit line where the *M*-unit at the top of the duct is the same as the *M*-unit at the bottom of the duct (illustrated by the dashed line).

Great semipermanent surface high-pressure systems, centered at approximately 30° north and south latitude, cover the ocean areas of the world. Poleward of these systems lay the mid-latitude westerly winds, and equatorward lay the tropical easterlies or the trade winds. Within these high-pressure systems, large-scale subsidence of air causes heating as the air undergoes compression. This leads to a layer of warm, dry air overlaying a cool, moist layer of air (often called the *marine boundary layer*).

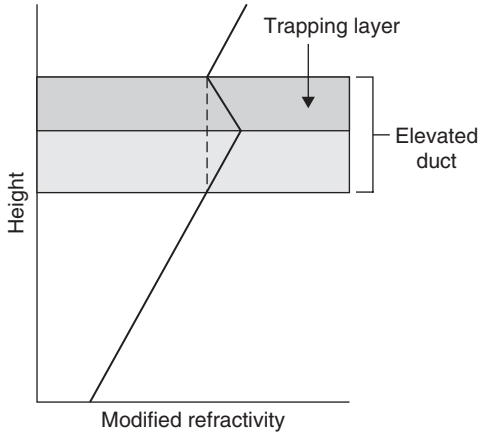


FIGURE 26.7 Elevated duct

The resultant inversion is referred to as the tradewind inversion and may create a strong ducting condition at the top of the marine boundary layer. Elevated ducts may vary from a few hundred meters above the surface at the eastern part of the tropical oceans to several thousand meters at the western part. For example, along the southern California coast, elevated ducts occur an average of 40% of the time, with an average top elevation of 600 meters. Along the coast of Japan, elevated ducts occur an average of 10% of the time, with an average top elevation of 1500 meters.

It should be noted that the meteorological conditions necessary for a surface-based duct are the same as those for an elevated duct. In fact, a surface-based duct may slope upward to become an elevated duct as warm, dry continental air glides over cool, moist marine air. The tradewind inversion may also intensify, thereby turning an elevated duct into a surface-based duct.

26.6 PROPAGATION MODELING[2,7]

Radio wave modeling is important for a number of reasons, all of which could be summarized into two large categories of engineering studies and operational performance. For engineering studies, the effects of propagation may be considered in new system design or in the evaluation of long-term performance of existing systems. For operational performance, consideration of propagation effects is usually based upon a single measured or forecasted atmosphere such that these effects can be exploited or mitigated by altering the system's use tactics. Over the years, many propagation models have been developed to account for the effects important to a particular application. These models span the spectrum of very fast executing but with low fidelity (simplified modeling of or the complete ignoring of certain propagation mechanisms) to relatively slow executing but with high fidelity (physically rigorous modeling of and full inclusion of all propagation mechanisms).

Spherical Spreading or Free-space Propagation Model. The simplest propagation model is spherical spreading where the transmitter and receiver are far removed

from the Earth's surface and the atmosphere, i.e., free space. Free space is defined as a region whose properties are isotropic, homogeneous, and loss-free. Spherical spreading models only consider the increasing surface area of a sphere centered on the transmitter and radiating out uniformly in all directions. The field strength at any point is inversely proportional to the square of the range between transmitter and the point. This is called *free-space path loss*. The power density, P_a, over a sphere at any point in free space is

$$P_a = \left(\frac{P_t G_t}{4\pi r^2}\right)(\mathrm{W/m^2}) \tag{26.8}$$

where P_t is the power radiated by the transmitter, r is the radius of the sphere, and G_t is the transmitting antenna's gain. For a loss-free, isotropic antenna, the gain is unity.

In free space, the power density at a loss-free, isotropic receiving antenna is the power density over the entire sphere's surface times the area of the sphere covered by the receiver antenna, also called the antenna's effective aperture, A_e. The effective aperture is related to the wavelength (λ) of radiation by

$$A_e = \frac{G\lambda^2}{4\pi} \tag{26.9}$$

Thus, the power at the receiver, P_r, for isotropic radiating and receiving antennas ($\boldsymbol{G_t} = \boldsymbol{G_r} = 1$) is

$$P_r = P_a A_e = \frac{P_t \lambda^2}{(4\pi r)^2} \tag{26.10}$$

The free-space path loss, L_{fs}, expressed in terms of the sphere's radius, r, and wavelength, λ, where r and λ are in the same units is

$$L_{\mathrm{fs}} = 10\log_{10}\left(\frac{P_t}{P_r}\right) = 10\log_{10}\left[\frac{(4\pi r)^2}{\lambda^2}\right] \tag{26.11}$$

Free-space path loss, L_{fs}, expressed in terms of range and frequency (in decibels) is given by

$$L_{\mathrm{fs}} = 32.45 + 20\log(f) + 20\log(r) \tag{26.12}$$

where f is frequency in MHz and r is the distance between the transmitter and receiver in kilometers. Free space is included in many modeling applications as a reference for other propagation effects.

If nonisotropic antenna radiational patterns are considered within the loss calculations, the loss is referred to as a propagation loss rather than a path loss. The propagation loss can be described with the aid of the propagation factor (F), which is defined as the ratio of the actual field strength at a point in space to the field strength that would exist at the same range under free-space conditions, with the beam of the transmitter directed toward the point in question. Symbolically this is

$$F = \frac{|E|}{|E_o|} \tag{26.13}$$

where E_o is the magnitude of the electric field under free-space conditions and E is the magnitude of the field to be investigated at the same point.

The propagation factor is a desirable quantity since it is an identifiable parameter in most radar equations. As stated earlier, it also contains all the information necessary to account for natural environmental effects. Thus, if the functional form of F is known, then the propagation loss at any point can be determined because the calculation of the free-space field is quite simple. The propagation loss, including antenna parameters, is equivalent to

$$L = L_{\text{fs}} - 20\log_{10}(F) \tag{26.14}$$

Effective-earth-radius Model. Since the majority of human activity takes place within the Earth's atmosphere, the free-space propagation model is usually inadequate for propagation assessment applications, and other propagation mechanisms need consideration. Under standard or normal atmospheric conditions, the radio ray curves downward with a curvature less than the Earth's surface. The effective-earth-radius concept[8] replaces the Earth's true radius with a larger radius such that the relative curvature between the ray and the Earth's surface is maintained, and the ray becomes a straight line. The effective-earth-radius, a_e, and the actual-earth-radius, a, are related by an effective-earth-radius factor, k, such that

$$a_e = ka \tag{26.15}$$

k may be computed using

$$k = \frac{1}{[1 + a(dn/dh)]} \tag{26.16}$$

where dn/dh is the vertical refractive index gradient. Using the mean-earth-radius of 6371 kilometers and a refractivity gradient of –39 N/km gives a k of 1.33 or about 4/3.

In addition to the consideration of refraction, other standard propagation mechanisms such as multipath interference, diffraction, troposcatter, and terrain may be included. Among the general class of effective-earth-radius models are the Standard Propagation Model (F-Factor),[2] the Terrain Integrated Rough Earth Model (TIREM),[8] the Irregular Terrain Model (ITM) also known as Longley-Rice,[9] and the Spherical Earth Knife Edge (SEKE).[10]

While these effective-earth-radius models are of the same nature, they do not implement the various propagation mechanisms equally. For example, the F-factor model properly implements multipath propagation for over-water surfaces whereas the TIREM model is based upon knife edge diffraction techniques, making the TIREM model inappropriate for over-water applications. Again, while troposcatter may be unimportant for active radar applications, its effects need to be included for applications of radar intercept by other sensors.

Waveguide Models. As engineering requirements demanded greater and greater fidelity, other modeling techniques were developed. One such modeling technique is using normal mode theory to compute field strength under standard or nonstandard refractive conditions. This class of models is referred to as waveguide models. The use of waveguide models dates back to the early 1900s when they were used to explain the propagation of long wavelength radio waves around the surface of the Earth in a waveguide formed by the Earth and the ionosphere. A description of waveguide models is beyond the scope of this chapter but may be found in a publication by Budden.[11]

While waveguide models are most useful for conditions where the vertical refractivity profile does not change along the propagation path (homogeneous environments), they can be applied to inhomogeneous environments by breaking the waveguide up into slabs in a technique known as *mode conversion*. While successful, this technique is less computationally efficient than other modeling techniques, and hence, waveguide models are not generally used for assessment systems requiring rapid execution times. Waveguide models serve as "laboratory benchmark" models against which the results of other modeling techniques can be compared. One such waveguide model is the MLAYER, derived from the original works of Baumgartner.[12]

Parabolic Equation Models. In 1946, Fock[13] used the parabolic equation (PE) method to describe electromagnetic propagation in a vertically stratified troposphere. In 1973, Hardin and Tappert[14] developed an efficient practical solution called the split-step Fourier method based upon fast Fourier transforms (FFTs) that has been widely applied to ocean acoustic propagation problems. The PE method and its solution by the split-step Fourier technique provides a very robust model for complicated refractivity structures; for within, near, and beyond the horizon effects; and is particularly good for propagation over irregular terrain. Thus, PE models allow for a single model assessment in many important applications. Three such PE models are the Terrain Parabolic Equation Model (TPEM),[15] the Tropospheric Electromagnetic Parabolic Equation Routine (TEMPER),[16] and the Variable Terrain Radio Parabolic Equation (VTRPE).[17]

Hybrid Models. While PE models are very attractive, they also have their disadvantages. Probably the biggest disadvantage is that they require very large computer resources, both in terms of memory and execution times, particularly for applications involving combinations of high frequencies, high elevation angles, high terminals, and long ranges. In some cases, this computational burden can be reduced by combining the best features of the various other models in a hybrid model. Once such model is the Advanced Propagation Model (APM) described by Barrios.[18] In APM, the PE model is combined with various ray optics and other phenomena models to create a hybrid model that can be up to 100 times faster than a PE model for stressful cases. Three other hybrid models are Radio Physical Optics (RPO),[19] TERPEM authored by Signal Science Limited,[20] and a hybrid method for computing transmission losses in an inhomogeneous atmosphere over irregular terrain by Marcus.[21]

Within APM, the assessment space is divided into four regions, or submodels, as illustrated in Figure 26.8. At ranges less than 2500 m and for all elevation angles above 5°, APM uses a flat earth (FE) model that ignores refraction and earth curvature effects. For ranges beyond the FE region where the grazing angles of reflected rays from the transmitter exceed a small limiting value, a full ray optic (RO) model is used that accounts for the effects of refraction and earth curvature. The PE model is used for ranges beyond the RO region, but only for altitudes below a maximum PE altitude determined by a maximum fast Fourier transform (FFT) size allowed. For ranges beyond the RO region and above the PE region, an extended optics (XO) method is used that is initialized by the PE model at the maximum PE altitude and uses ray optics methods to propagate the signal to higher altitudes. Continuity of the solutions across each region's boundaries is kept less than 0.1 dB by careful selection of the limiting RO grazing angle and the maximum PE propagation angle.

The propagation models within APM have also been combined with other environmental effects models, such as gaseous absorption and surface clutter, to form a complete propagation package. The physical propagation phenomena considered by APM version 2.0.01 are illustrated in Table 26.2. As can be seen from the table,

TABLE 26.2 Propagation Effects Modeled by APM

Propagation Effects Mechanisms	Due To Environment			APM 2.0.01
	Sea	Terrain	Atmosphere	
Range-dependent refractive conditions			●	✓
Variable terrain		●		✓
Multi-path	●	●	●	✓
Diffraction	●	●		✓
Terrain masking		●		✓
Troposcatter			●	✓
Rough (Sea) surface	●			✓
HF surface wave	●			✓
Range-dependent dielectrics	●	●		✓
Obstacle gain		●		✓
Surface clutter	●	●		✓
Gaseous abosrption			●	✓
Rain attenuation			●	✗
Vegetation		●		✗

APM considers almost every environmental effect, making it a highly desirable model for use in complex assessment systems.

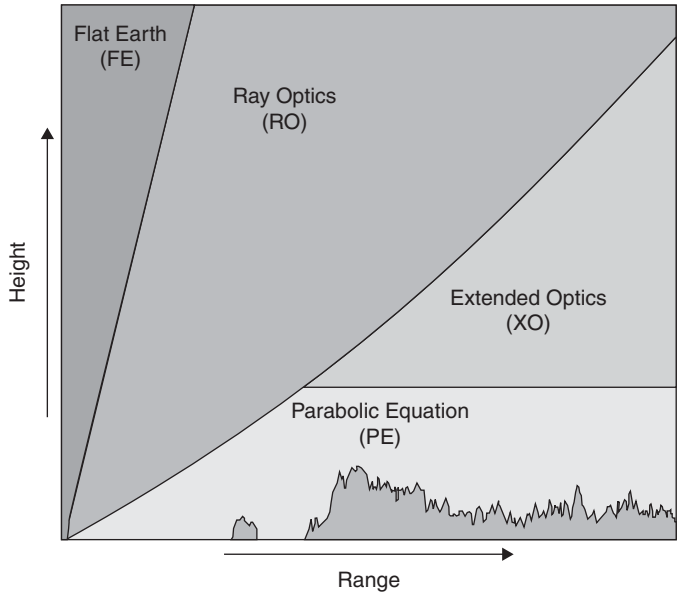


FIGURE 26.8 APM submodel regions

26.7 EM SYSTEM ASSESSMENT PROGRAMS

Using the power of the personal computer in conjunction with the maturity of EM system and environmental propagation modeling, assessment programs and associated software allow a user to define and manipulate refractivity and other natural environment data, run propagation models on that data, and display the results in terms of expected performance on actual or proposed electromagnetic systems. While there are several assessment systems in use within the U.S. and in various other countries, the following discussion is limited to AREPS. It is used extensively throughout the U.S. Department of Defense (DoD) and other federal government agencies, U.S. private industry, their foreign counterparts, and by private individuals.

AREPS grew out of an urgent military operational requirement for radar performance and propagation modeling within a terrain effects dominated environment. The assessment system requirements included modeling all natural environmental effects, being quickly executable, and executing on a Microsoft Windows operating system personal computer. An evaluation of the various propagation models for their strengths and weaknesses quickly showed that a hybrid model was the only acceptable solution. The AREPS graphical user interface was created and interfaced to the APM to provide the user an end-to-end radar propagation assessment tool. Because EM propagation effects are not just limited to radar frequencies, over time the initial radar requirements for AREPS were expanded beyond simple radar detection applications to include applications in communications and electronic warfare. AREPS is the only approved EM system assessment application within the Department of the Navy Chief Information Officer DON Applications & Database Management System (DADMS). APM is the only accredited (by the Chief of Naval Operations) EM propagation (2 MHz to 57 GHz) model for use in Navy systems. Both AREPS and APM are accredited within the Navy Modeling and Simulation Office (NMSO). AREPS is also a North Atlantic Treaty Organization (NATO) application, approved by the Military Committee Meteorological Group/Working Group, Battle Area Meteorological Systems and Support plus with Partners.

AREPS version 3.6 contains several EM propagation models for applications at various frequencies. For frequencies of 2 MHz to 57 GHz, AREPS uses the APM. For HF sky wave communications, AREPS uses an HF modeling suite,[22] consisting of a fully 3D ionosphere ray trace model, an HF field strength model, and an HF noise model. In addition to these EM propagation models, AREPS may optionally use two internationally recognized ionosphere models, the Parameterized Ionospheric Model (PIM)[23] and the International Reference Ionosphere (IRI).[24] In addition to the propagation models, AREPS contains a system performance radar model, which is discussed in the next section.

AREPS considers range- and azimuth angle–dependent influences from surface features to include terrain elevation, finite conductivity, dielectric ground constants, and scattering effectiveness factors. The terrain elevation data may be obtained from the National Geospatial-Intelligence Agency's (NGA) Digital Terrain Elevation Data (DTED) or from any other suitable source. The finite conductivity and dielectric ground constants may be selected from those defined by the International Telecommunication Union, International Radio Consultative Committee (CCIR),[25] or from any other suitable source.

AREPS considers range- and azimuth angle–dependent atmospheric refractivity data derived from the upper-air observations of radiosondes, other sensors, or mesoscale meteorological models such as the U.S. Navy Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS).[26] Radiosonde data may be manually entered or automatically decoded from either the World Meteorological Organization (WMO) observational message format or a free-form column format, obtained from a number of different data sources. In addition, climatological refractive conditions may be selected from a 921 WMO station, worldwide reporting database. For ocean reporting stations or numerical weather prediction grid points over the ocean, AREPS automatically calculates an evaporation duct refractive profile and appends it to the bottom of the upper-air observation for a complete description of the propagation environment.

AREPS computes and displays a number of EM system performance assessment tactical decision aids. These are radar probability of detection, ESM vulnerability, LF to EHF communications, simultaneous radar detection and ESM vulnerability, and surface-search detection ranges. All decision aids are displayed as functions of range, azimuth angle, and/or height. Detection probability, ESM vulnerability, and communications assessments are based on EM system parameters stored in a user-defined and changeable database. In addition to normal radar parameters, a user may completely define the antenna radiation pattern to account for sidelobe considerations. The database also includes radar target descriptions and platforms' EM emitter suites.

Figure 26.9 is an illustration of an AREPS four-panel display. This display was created in support of U.S. homeland defense for the February 5, 2006, Super Bowl XL. The Super Bowl XL (Roman numeral for 40) is a championship football game

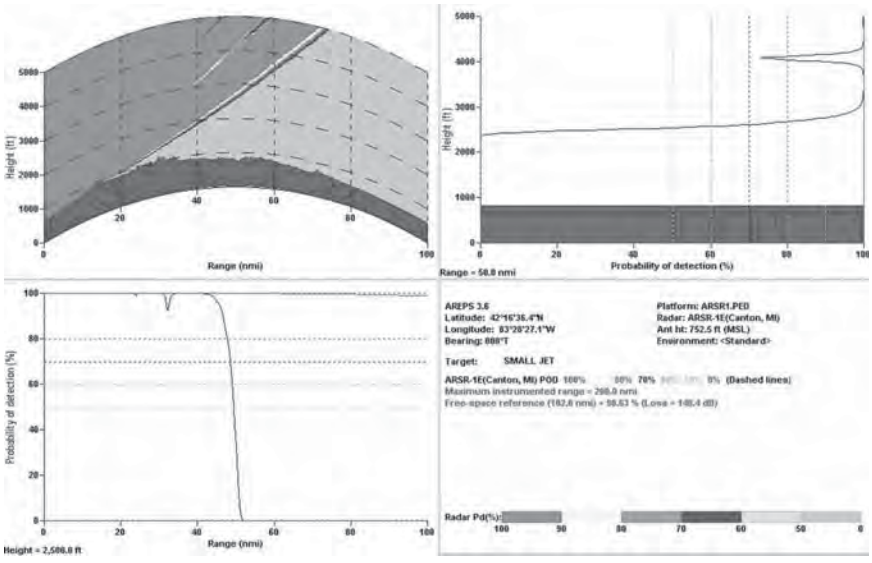


FIGURE 26.9 AREPS homeland defense application

played in the United States. It was attended by over 60,000 people, all confined in one sporting event stadium, which presented a significant target for a terrorism attack. The display shows a radar probability of detection in a range versus height depiction, radar probability of detection versus range at a constant altitude depiction, and radar probability of detection versus height at a constant range depiction. This display is courtesy of the 84th Radar Evaluation Squadron, Hill Air Force Base, Utah. The radar depicted is the ARSR-1E (U.S. Air Route Surveillance Radar) located at Canton, Michigan. The target of interest is a small private aircraft.

As the personal computer version of AREPS developed, it also transitioned into an application called the Naval Integrated Tactical Environmental Subsystem (NITES), a segment of the Global Command and Control System - Maritime (GCCS-M). In NITES, the AREPS functionality of the personal computer was coded in Java and interfaced to the Common Operating Picture (COP). The COP is a real-time display of tactical information and current force positioning. Thus, a radar assessment provided by AREPS may be displayed as a tactical overlay upon the operating picture. Figure 26.10 is an illustration of such a COP display. In this illustration, the radar coverage of three coastal surveillance radars is shown along with the traditional height versus range coverage display typical of the personal computer version of AREPS.

As the Navy progressed into web-based applications, the display of AREPS tactical decision aids followed suit. One such web-based application is the Navy's Composable ForceNet, an application similar in nature to the GCCS-M. Figure 26.11 is an illustration of the Composable ForceNet application. For this illustration, the background for the Composable ForceNet COP is a hypothetical ocean and island operation area. The symbology (such as the small circles, squares, and half circles), represent the disposition of various forces such as ships and, aircraft. The multiple elliptical and fan-shaped shaded areas correspond to a certain radar probability of detection of various targets by various operational radars. For example, the small fan-shaped shaded area in the upper-right corner of the display represents the area where a ground-based radar would be able to detect (with a certain probability of detection) a particular aircraft target.

An optional capability of AREPS is to provide to an external application, the computed radar probability of detection, propagation loss, or propagation factor, to an external application. Thus, a tactical application developer need not have knowledge of the underlying radar propagation modeling techniques or other environmental considerations but can rely upon AREPS to "serve up" data for display within his or her own application. Such an application is the Simulation and Display System (SIMDIS™), authored by the Naval Research Laboratory, Washington DC.[27] Figure 26.12 is an illustration of AREPS-computed data displayed three-dimensionally within SIMDIS™. In this figure, the view is looking northward within the byte of Southern California. The eastern terrain of San Clemente Island shows in the middle left of the figure. From this current view, a ship with a particular air-search radar is located beyond (north) of the island. The darkened fan-shaped area extending upward over the island represents the radar's probability of detection of a certain target as the radar sweeps southward over the island. For a target located in the forefront of the island and at a sufficiently low altitude, the display shows the target is not detectable by the radar as the target is masked by the intervening terrain.

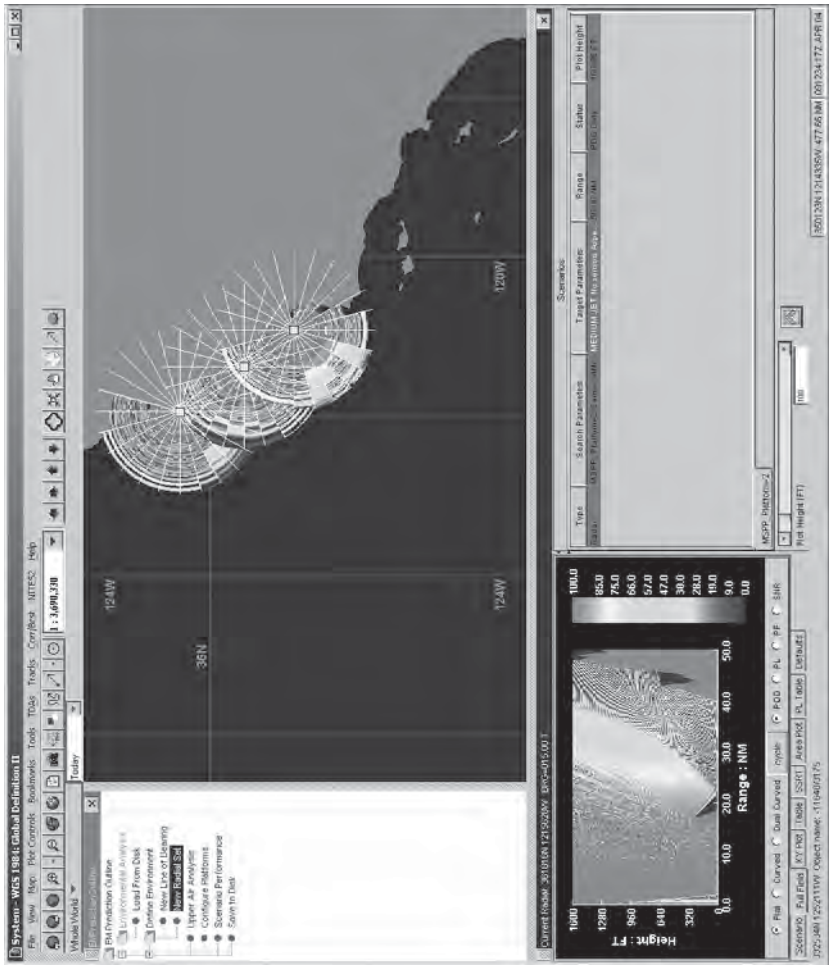


FIGURE 26.10 AREPS display upon the COP within NITES

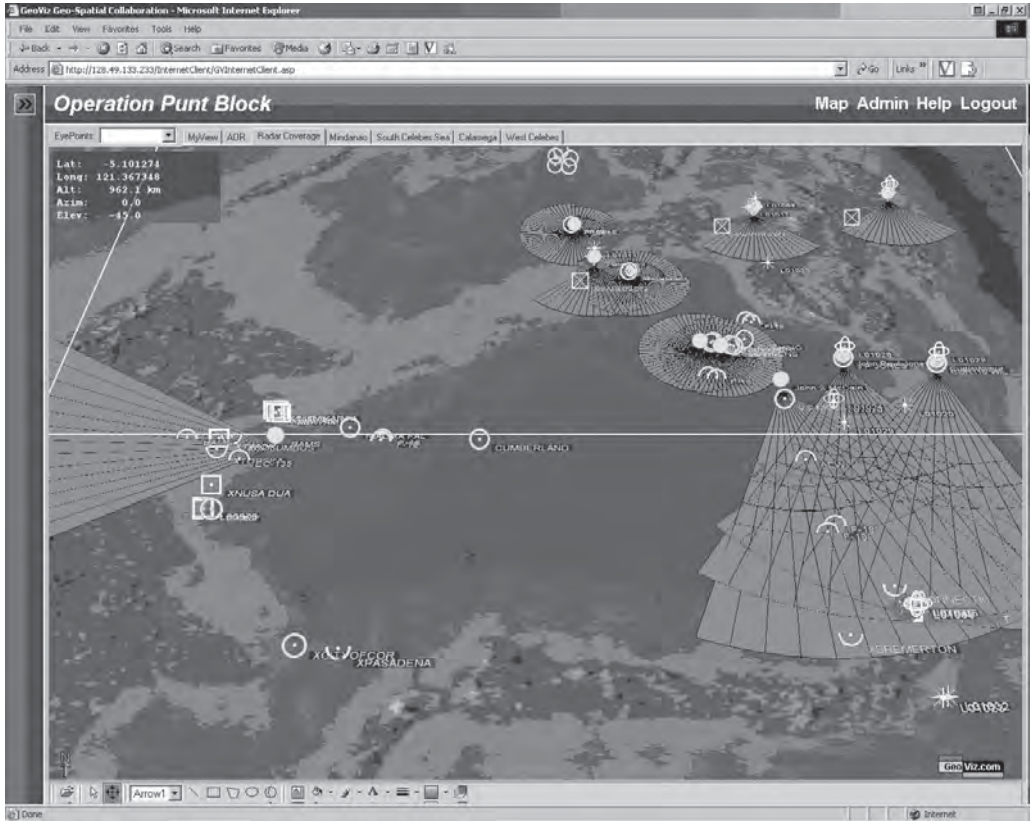


FIGURE 26.11 AREPS computations display upon the Composable ForceNet COP

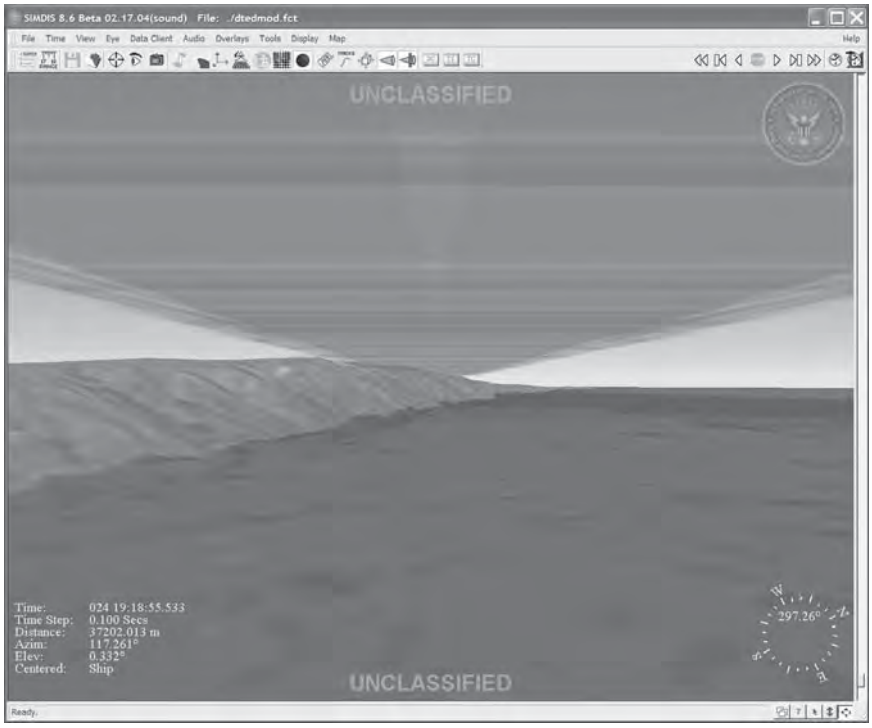


FIGURE 26.12 AREPS computations display within SIMDIS showing radar probability of detection of a small target under the influence of terrain-masking effects

26.8 AREPS RADAR SYSTEM ASSESSMENT MODEL

The main purpose of AREPS is to offer radar operators the ability to visualize their radar's detection of threat targets or their own platform's detection by threat radars under various natural environmental conditions. The primary visualization is a height versus range display of radar probability of detection, where the radar probability of detection expressed in percent corresponds to certain energy levels in height and range from the radar. APM computes propagation loss in dB. To make a determination of radar performance, AREPS needs to make a comparison of free-space propagation loss and propagation loss within the Earth's environment as computed by APM. Thus, AREPS contains a fairly simplistic pulsed-radar model to calculate free-space propagation loss from radar system parameters such as frequency, pulse length, etc. The models to do this calculation are taken from Blake[28] and are fully described within the AREPS online help and the AREPS operator's manual. Thus, AREPS' radar system assessment model will not be repeated here. The intent of this section is to show the necessary radar system and radar target inputs needed by the AREPS program. These inputs are shown in Figures 26.13 and 26.14. For a complete description of each input parameter, you may refer to the AREPS online help or the AREPS operator's manual.

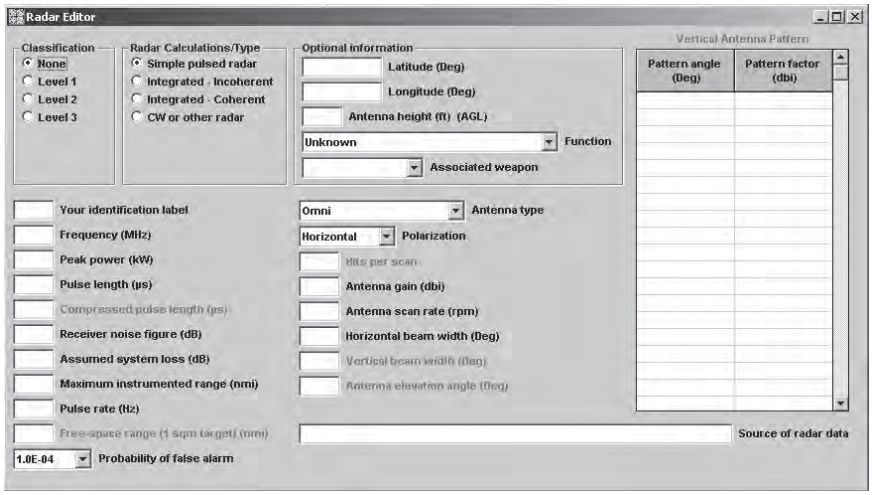


FIGURE 26.13 AREPS radar window

To assist the AREPS user, these and other input windows have many "canned" default values combined with many unit options. One example is the transmitting antenna pattern. APM will consider the full antenna pattern of the transmitting antenna.

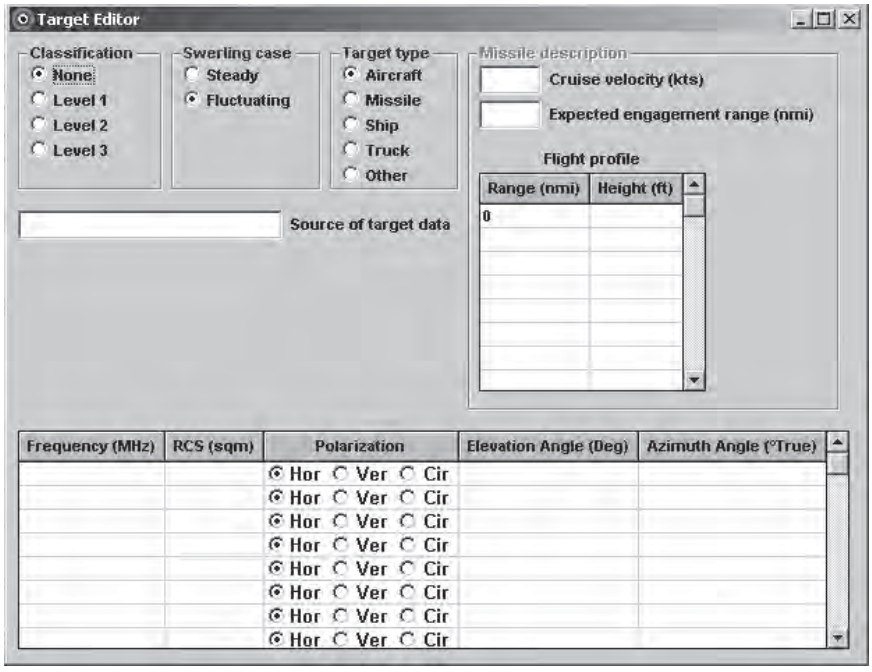


FIGURE 26.14 AREPS target window

AREPS provides some basic "canned" antenna patterns such as omni, SinX/X, cosecant-squared, and a generic height-finder. With a user-specified antenna type, the operator may enter the pattern angle and factor directly. In addition to entering the antenna pattern directly from the keyboard, AREPS also provides the capability of importing an antenna pattern from an ASCII text file that you may have created from another application. An example of multiple units is the transmitter's peak power. The default unit for peak power is Kilowatts. By right-clicking on the label associated with peak power, you may select other units of input. AREPS will even convert the input number automatically from one unit to the other unit.

26.9 AREPS RADAR DISPLAYS

By default, the results of the propagation model calculations are shown in terms appropriate for a tactical radar operator, i.e., a height versus range display of radar probability of detection as a percentage, as illustrated in Figure 26.15. This display is referred to as a tactical decision aid because it will allow the radar operator to make some sort of tactical decision. For example, superimposed on this tactical decision aid is the flight profile of a target missile (the solid line sloping downward to the origin from right to left). The atmospheric environment is surface-based ducting. The nondetection skip-zone is clearly seen in addition to the target's flight pattern, allowing the operator to see how the probability of detection of the target will vary with range and height. With this knowledge, a decision about when to attempt an attack on the missile can be made.

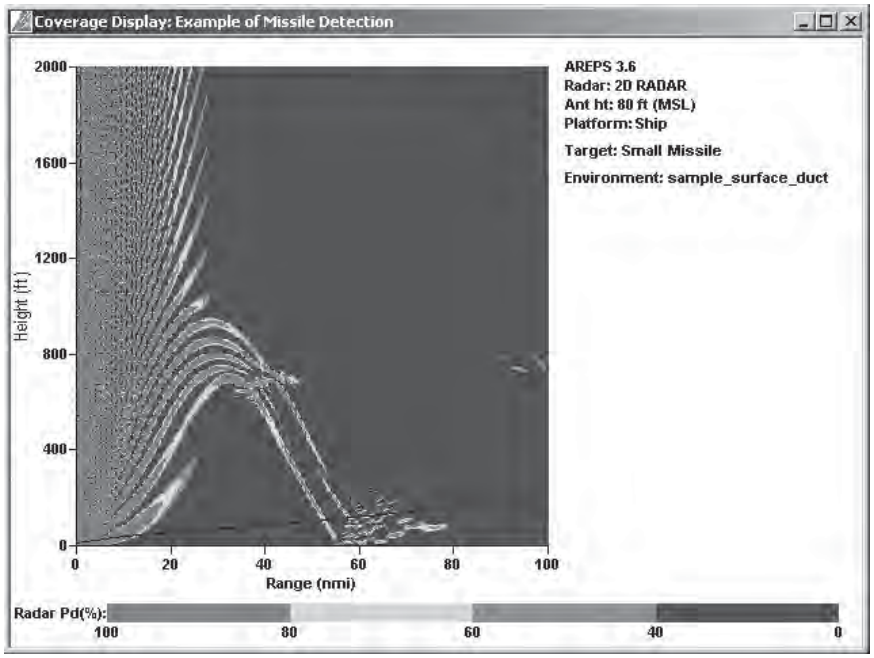


FIGURE 26.15 AREPS height versus range coverage for surface-search radar and small missile target—probability of detection

For radar engineers, a display such as that shown in Figure 26.15 will most likely not be very useful since propagation factor, F_p, is the desired quantity for the radar range equation and not the propagation loss as computed by APM. However, there is a simple relationship between propagation loss and propagation factor. This is

$$F_p = L_{\text{fs}} - L_{\text{dB}} \tag{26.17}$$

where L_{fs} is the free-space propagation loss given by Eq. 26.12 and L_{dB} is the propagation loss in dB as computed by APM. As a convenience for the radar engineer, AREPS will also display the APM output in terms of propagation factor. Thus, Figure 26.15 displayed in terms of radar probability of detection as a percentage will appear as shown in Figure 26.16 when displayed in terms of propagation factor.

In addition to the default height versus range display, AREPS contains many other display and data output options. For example, Figure 26.17 shows (for the same missile detection example) a signal-to-noise ratio versus range display superimposed with a clutter-to-noise ratio computed from an 8 meter-per-second wind speed. The display altitude is 100 feet above sea-level. One can see the effects of multipath interference between a range of 0 and about 20 nautical miles. Beyond the multipath interference region, one can see a major fall off of signal-to-noise and clutter-to-noise ratio comes within the surface-based duct skip-zone between about 20 nautical miles and 50 nautical miles. Note, however, the increase in ratios for ranges beyond the skip-zone. In fact, it is even seen that for ranges between 55 and 65 nautical miles, the clutter-to-noise ratio exceeds the signal-to-noise ratio. Thus for these ranges, the radar return is overwhelmed with clutter.

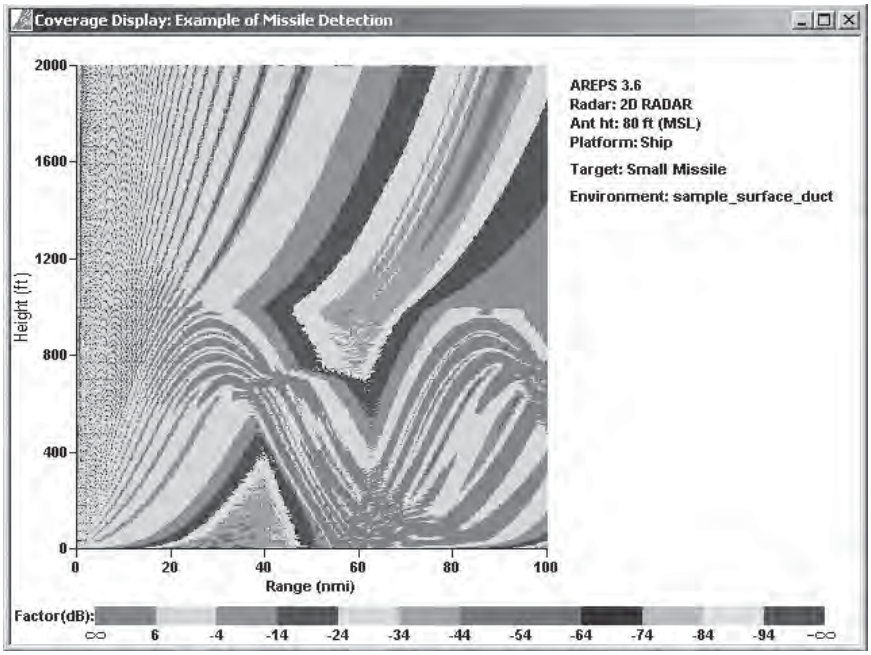


FIGURE 26.16 AREPS height versus range coverage for surface-search radar and small missile target—propagation factor

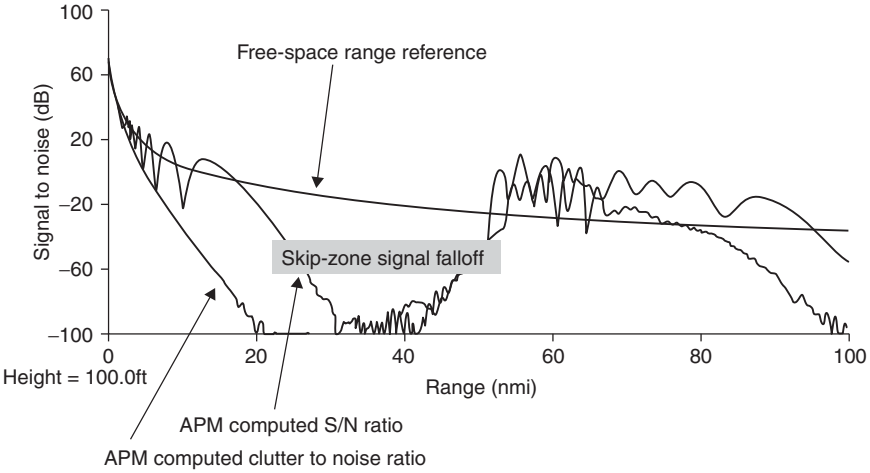


FIGURE 26.17 AREPS signal-to-noise ratio versus range for surface-search radar and small missile target—clutter-to-noise ratio superimposed

While a graphic may be useful for visual inspections, it may not be useful for an engineering analysis task. AREPS has many other data display features. For example, the propagation factor values as shown in Figure 26.16 may also be exported in a number of different text formats for use in other engineering applications.

REFERENCES

1. W. L. Patterson, "Advanced refractive effects prediction system," Space and Naval Warfare Systems Center TD 3101, January 2000. AREPS may be freely obtained at http://areps .spawar.navy.mil.
2. W. L. Patterson et al., "Engineer's Refractive Effects Prediction System (EREPS)," Naval Command, Control, and Ocean Surveillance Center TD 2648, May 1994.
3. E. Brookner, "Radar performance during propagation fades in the Mid-Atlantic region," *IEEE Transactions on Antennas and Propagation*, vol. 46, No. 7, July 1998.
4. M. P. M. Hall, "Effects of the troposphere on radio communications," London: Institution of Electrical Engineers, 1979, p. 30.
5. K. D. Anderson, "Radar detection of low-altitude targets in a maritime environment," *IEEE Transactions on Antennas and Propagation*, vol. 43. no. 6, June 1995.
6. S. E. Morison, "Aleutians, Gilberts and Marshalls, June 1942–April 1944. History of United States Naval Operations in World War II, Volume VII," *Military Affairs*, vol. 15, no. 4, pp. 217–218, 1951.
7. H. V. Hitney, "Refractive effects from VHF to EHF, part B: propagation models," Advisory Group for Aerospace Research & Development, AGARD-LS-196, pp. 4A1–4A13, September 1994.
8. J. R. Powell, "Terrain Integrated Rough Earth Model (TIREM)," Rep. TN-83-002, Electromagnetic Compatibility Analysis Center, Annapolis, MD, September 1983.
9. A. G. Longley and P. L. Rice, "Predictions of troposphere radio transmission loss over irregular terrain: A computer method," Environmental Science Services Administration Tech. Rep. ERL 70-ITS 76, U.S. Govt. Printing Office, Washington, DC, 1968.
10. S. Ayasli, "SEKE: A computer model for low altitude radar propagation over irregular terrain," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 8, August 1986.

11. K. G. Budden, *The Wave-Guide Mode Theory of Wave Propagation*, Inglewood Cliffs, NJ: Prentice-Hall, Inc., 1961. Also London: Logos Press, 1961.
12. G. B. Baumgartner, "XWVG: A waveguide program for trilinear tropospheric ducts," Naval Ocean Systems Center TD 610, June 1983.
13. V. A. Fock, *Electromagnetic Diffraction and Propagation Problems*, New York: Pergamon, 1965.
14. R. H. Hardin and F. D. Tappert, "Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev., 15, 2, p. 423, 1972.
15. A. E. Barrios, "A Terrain Parabolic Equation Model for Propagation in the Troposphere," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 1, pp.90–98, January 1994.
16. G. D. Dockery, "Modeling electromagnetic wave propagation in troposphere using the parabolic equation," *IEEE Transactions on Antennas and Propagation*, vol. 36, pp. 1464–1470, October 1988.
17. F. J. Ryan, "Analysis of electromagnetic propagation over variable terrain using the parabolic wave equation," Naval Ocean Systems Center TR-1453, October 1991.
18. A. E. Barrios, "Advanced Propagation Model (APM) Computer Software Configuration Item (CSCI)," Space and Naval Warfare Systems Center TD 3145, August 2002.
19. H. V. Hitney, "Hybrid ray optics and parabolic equation methods for radar propagation modeling," in *Radar 92*, IEE Conf. Pub. vol. 365, October 12–13, 1992, pp. 58–61.
20. Ken Craig and Mireille Levy, http://www.signalscience.com.
21. S. W. Markus, "A hybrid (Finite Difference-surface Green's Function) method for computing transmission losses in an inhomogeneous atmosphere over irregular terrain," *IEEE Transactions on Antennas and Propagation*, vol. 40. no.12, p. 1451–1458.
22. R. B. Rose, "Advanced prophet HF assessment system," Naval Ocean Systems Center, San Diego, January 1984.
23. R. E. Daniel, Jr., L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka, and R. W. Schunk, "Parameterized ionospheric model: A global ionospheric parameterization based on first principal models," *Radio Science*, vol. 30, pp. 1499–1510, 1995.
24. K. Rawer, S. Ramakrishnan, and D. Bilitza, "International reference ionosphere 1978," International Union of Radio Science, URSI Special Report, pp. 75, Bruxelles, Belgium, 1978.
25. "Propagation in Non-ionized Media," International Telecommunication Union, International Radio Consultative Committee (CCIR), vol. V, Report 879-1, p 82.
26. "U.S. Navy Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS)," Naval Research Laboratory, Marine Meteorology Division, NRL Publication 7500-03-448, May 2003.
27. Naval Research Laboratory, Washington, DC, https://simdis.nrl.navy.mil, simdis@enews.nrl.navy.mil.
28. L. V. Blake, *Radar Range Performance Analysis*, Lexington, MA: Lexington Books, D.C. Heath and Co., 1980.

INDEX

A
A-12 low cross section aircraft, **14.**40
Absorbers, **14.**32, **14.**36 to **14.**38
Active aperture, **13.**53 to **13.**55
Active Electronically Scanned Array (AESA), **5.**1, **5.**8 to **5.**10, **10.**28
  compared to mechanical scan, **5.**9 to **5.**10
  timing structure, **5.**15 to **5.**16
  typical waveform parameters, **5.**13, **5.**15
  waveform variations for air-to-surface mode, **5.**12 to **5.**13
Active jamming, **24.**5
Active-switch modulator, **10.**24
A/D converter
  and MTI dynamic range, **2.**78 to **2.**80
Adaptive arrays (ECCM), **24.**20 to **24.**30
Adaptive jammer and clutter cancellation, **24.**25
Adaptive MTI, **2.**80 to **2.**83
Adaptive thresholding, **7.**11 to **7.**19
Advanced Propagation Model (APM), **26.**16 to **26.**17
Advanced Refractive Effects Prediction System (AREPS), **26.**1 to **26.**2, **26.**18 to **26.**27
Aid to Navigation (AtoN), **22.**25 to **22.**26
Air-to-air ground ranging, **5.**33
Air-to-air mission profile, **5.**12 to **5.**14
Air-to-air modes, **5.**14, **5.**16 to **5.**28
  medium PRF, **5.**16 to **5.**20
Air-to-surface mission profile, **5.**10
Air-to-surface mode suite, **5.**11 to **5.**12
Air-to-surface radar modes, **5.**28 to **5.**42
Air-traffic control (ATC) radar, **1.**21 to **1.**22
Air Traffic Control Radar Beacon System (ATCRBS), **7.**49
Airborne Early Warning, **3.**1.
  See also Airborne Moving Target Indication
  antenna sidelobes, **3.**13 to **3.**14
  coverage, **3.**2 to **3.**3
  platform motion, **3.**3 to **3.**14
  TACCAR, **3.**4 to **3.**9
Airborne Moving Target Indication (AMTI)
  example of, **3.**32 to **3.**33
  improvement factor, **3.**8
  multiple spectra, **3.**31 to **3.**32
  scanning motion compensation, **3.**14 to **3.**17
  and STAP, **3.**23 to **3.**31
Aircraft radar cross section, **14.**13 to **14.**14
Alpha-beta (α-β) filter, **7.**26 to **7.**27, **7.**30
Altimeters, from space, **18.**33 to **18.**57
Altitude line, **4.**4 to **4.**5
Ambiguity function, **8.**40
Ambiguity resolution, in MTI, **2.**89 to **2.**91
Amplitude-comparison monopulse, **9.**3 to **9.**11
Amplitude noise, in tracking, **9.**27 to **9.**30
AN/APG-76, **17.**25
AN/APY-9, **3.**32 to **3.**33
AN/FPQ-6, **9.**2, **9.**9, **9.**18, **9.**21, **9.**44
AN/FPS-23, **23.**7
AN/EPS-117, **8.**30, **13.**63 to **13.**64
AN/MPS-39, **9.**2, **9.**26
AN/SPS-40, **10.**27, **11.**32 to **11.**33
AN/SPY-1, **13.**62

AN/SPY-3, **13.**69
AN/TPS-59, **8.**30, **13.**63 to **13.**64
AN/TPS-63, **12.**18
AN/TPS-77, **13.**63 to **13.**64
AN/TPS-78, **13.**65 to **13.**66
Analog-to-digital (A/D) converters, **6.**35 to **6.**40

Anechoic chamber, **14.**32
Angle accuracy, by beam splitting, **7.**5 to **7.**7
Angle measurement errors, **9.**43 to **9.**44
Angle noise, **9.**30 to **9.**35
Angle scintillation (glint), **9.**30 to **9.**35, **9.**47
Angular direction, **1.**8
Anomalous propagation, **26.**6 to **26.**13
Antenna-related ECCM, **24.**10 to **24.**31
Antennas, **1.**2, **1.**4 to **1.**5.
  See also phased arrays, reflector antennas
  basic principles, **12.**3 to **12.**15
  for civil marine radar, **22.**10 to **22.**12
  for ground penetrating radar, **21.**24 to **21.**30
  phased array, **13.**2 to **13.**3
  for pulse doppler, **4.**12 to **4.**13
  reflector antennas, role of, **12.**1
  reflector antennas, types of, **12.**7
Anti-radiation missile (ARM), **24.**40
Anti-range-gate pull-off (A-RGPO), **24.**46
Aperture gain, of antennas, **12.**4 to **12.**5
Apollo Lunar Sounder, **18.**60 to **18.**61
Aquarius scatterometer, **18.**58
Area MTI, **2.**84
Array feeds, for reflector antennas, **12.**28 to **12.**30
ARSR-4, **12.**30
ASR-9, **10.**27 to **10.**28, **12**.20 to **12.**21
ASR-11 MTI filter design, **2.**50 to **2.**51
ASR-12, **8.**30, **10.**27 to **10.**28
Atmosphere, of the Earth, **26.**2
Atmospheric ducts, **26.**7 to **26.**10
Attenuation, by clouds, **19.**7 to **19.**8
  by fog, **19.**12
  by hail, **19.**11 to **19.**12
  by rain, **19.**8 to **19.**11
Automatic detection, **7.**1 to **7.**2, **7.**20 to **7.**22
Automatic gain control (AGC), in monopulse tracker, **9.**5, **9.**10 to **9.**11
Automatic Identification System (AIS)
  in civil marine radar, **22.**23 to **22.**25
  in search and rescue, **22.**27
Automatic noise-level control, **6.**23 to **6.**24
Automatic Radar Plotting Aid (ARPA), **22.**17
Automatic tracking, **7.**22 to **7.**46
  alpha-beta (α-β) filter, **7.**26 to **7.**27, **7.**30
  detection acceptance, **7.**25 to **7.**26
  Interacting Multiple Model (IMM), **7.**35 to **7.**37
  Kalman filter, **7.**28 to **7.**35
  new track formation, **7.**41 to **7.**46
  retrospective processing, **7.**42 to **7.**43
  scheduling and control, **7.**46
  track association, **7.**38 to **7.**41
  track file, **7.**23 to **7.**25
  updating tracks, **7.**26 to **7.**30
AWACS, **13.**65

B
B-2 low cross section aircraft, **14.**18, **14.**42
Ballistic missile defense, **13.**54
Bandwidth
  importance of, **1.**8 to **1.**9
  of phased arrays, **13.**38 to **13.**45
  receiver, **6.**9
Bar, in airborne radar, **5.**15 to **5.**16
Barker codes, **8.**17

Batch processor, **7.**8 to **7.**11
Battle of the Pips, ducting effects, **26.**9
Beacon equation, **24.**4
Beacon rendezvous, **5.**27
Beam-shape factor, **7.**3
Beam-shape loss, **4.**40
Beam-splitting angle accuracy, **7.**5 to **7.**7
Beam steering, **13.**15 to **13.**19
Beam steering control, **5.**10
Binary integrator, **7.**7 to **7.**8, **7.**12 to **7.**13
Birds, and MTI, **2.**96 to **2.**98
Bistatic plane, **23.**3
Bistatic radar
  applications, **23.**9 to **23.**14
  benchmark range, **23.**4 to **23.**5
  concept of, **23.**1 to **23.**3
  coordinate system, **23.**3 to **23.**4
  doppler in, **23.**14 to **23.**17
  floodlight beams, **23.**27
  forward scatter, **23.**21
  glint in, **23.**20 to **23.**21
  ground echo, **16.**29
  hitchhiker, **23.**10
  with noncooperative transmitter, **23.**29 to **23.**31
  ovals of Cassini, **23.**6 to **23.**8
  passive, **23.**10
  pseudo-monostatic region, **23.**19 to **23.**20
  pulse chasing, **23.**28 to **23.**29

range equation, **23.**5 to **23.**6
SAR, **23.**17
scattering coefficient, **23.**22 to **23.**25
surface clutter in, **23.**22 to **23.**26
target cross section, **23.**19 to **23.**21
target location, **23.**17 to **23.**19
with TV transmitter, **23.**29 to **23.**31
Blind speeds, **2.**9 to **2.**10
Block diagram
digital receiver, **25.**2
medium PRF air-to-air pulse doppler, **5.**17
monopulse tracking radar, **9.**4 to **9.**5
Moving Target Detection, **2.**6 to **2.**7
multifunctional fighter aircraft radar, **5.**3
pulse compression, **8.**1 to **8.**2
pulse doppler, **4.**11
radar, **1.**3
SBR altimeter, **18.**40
Bodies of revolution radar cross section, **14.**39 to **14.**40
Bragg scatter, **15.**2, **15.**28 to **15.**32, **15.**38,
and ground echo, **16.**11 to **16.**12
Burst mode, in SBR SAR, **18.**23 to **18.**24

C
C band, **1.**16
Calibrate and self-test, **5.**42
Calibration of meteorological radar, **19.**18 to **19.**19
Cascaded Integrator-Comb (CIC) Filters, **25.**29 to **25.**32
Cassegrain antenna, **12.**21 to **12.**23, **12.**25, **14.**34
Cassini, **18.**46
Cathode pulser, **10.**23
Cell-averaging CFAR, **7.**11 to **7.**12, **7.**17
CFAR, **7.**11 to **7.**18
and ECCM, **24.**35 to **24.**36
loss, **4.**44
probability of detection, **7.**15
target suppression in, **7.**16 to **7.**17
Chaff, **24.**5 to **24.**6
Chaff and ECCM, **24.**19, **24.**34
Chart radars, **22.**22 to **22.**23
Chebyshev filter bank, **2.**54 to **2.**55
Civil Marine Radar (CMR), **22.**33
and aids to navigation (AtoN), **22.**25, **22.**26
antennas, **22.**10 to **22.**12
Automatic Identification System (AIS), **22.**2, **22.**8
integration with, **22.**23 to **22.**25
cost of, **22.**3
detection performance, **22.**4 to **22.**6
detection and processing, **22.**13 to **22.**15
display, **22.**21
early days of, **22.**31 to **22.**33
environmental conditions, **22.**3
international standards for, **22.**7 to **22.**10
magnetrons for, **10.**16
precipitation and sea clutter, **22.**5
racons and, **22.**26
radar beacons and, **22.**25
radar target enhancers (RTEs), **22.**27
RF head (transmitter and receiver), **22.**12 to **22.**13
search and rescue transponders, **22.**27
solid-state, **22.**16 to **22.**17
target tracking, **22.**17 to **22.**19
user interface, **22.**19 to **22.**22

validation testing of, **22.**28 to **22.**29
vertical lobing in, **22.**6 to **22.**7
Class of (amplifier) operation, **11.**18 to **11.**20
Clouds, attenuation in, **19.**7 to **19.**8
CloudSat, **18.**63, **19.**39
Clustered-cavity klystron, **10.**12 to **10.**13
Clutter
amplitude characteristics, **2.**17
attenuation, in MTI, **2.**20 to **2.**21
characteristics for MTI, **2.**10 to **2.**19
exponential model, **2.**12 to **2.**16
in ground penetrating radar, **21.**5, **21.**10
maps, **2.**83 to **2.**87, **6.**23, **7.**19
models, for ground echo, **16.**29 to **16.**34
MTI filter design, **2.**33 to **2.**46
optimum filter design, **2.**25 to **2.**33
point scatterers, **2.**18 to **2.**19
in pulse doppler radar, **4.**14 to **4.**24
reflectivity, **2.**17 to **2.**18
Clutter-limited detection in pulse doppler, **4.**48
Coaxial magnetron, **10.**14 to **10.**15
Coaxitron, **10.**21
Cobra Dane radar, **8.**36
Coherent processing, and ECCM, **24.**34 to **24.**35
Coherent Processing Interval (CPI), **2.**7
Coherent on receive, **10.**14

COHO, **6.**3, **6.**20
Collapsing loss, **7.**3
Compact range, **14.**32 to **14.**33
Complementary pulse compression waveforms, **8.**19
Complex envelope representation, **8.**38 to **8.**39
Computer codes for reflector design, **12.**33 to **12.**35
Conformal arrays, **13.**3
Conical-scan tracking, **9.**16 to **9.**17
Conopulse, **9.**15
Constant false alarm rate. *See* CFAR
Continuous wave (CW) radar, **1.**5
Constant Efficiency Amplifier (CEA), **10.**21 to **10.**22, **10.**26
Cooperative target identification, **5.**22 to **5.**23
CORDIC processor, **25.**22 to **25.**25
Corner reflector, radar cross section, **14.**9 to **14.**11
Corporate feed monopulse phased array, **9.**14
Co secant-squared antenna, and STC, **2.**98
COSMO-SkyMed SAR, **18.**12
Costas codes, **8.**25 to **8.**26
Counter Battery Radar (COBRA), **13.**62
Creeping waves, **14.**3, **14.**5
Critical frequency, in HE OTH, **20.**3
Crossed-field amplifier (CFA), **10.**3, **10.**16 to **10.**17
Crossed-field tubes, **10.**2
Cross-eye ECM, **24.**42 to **24.**43
Cross-polarization jamming, **24.**43
Cross-polarization tracking, **9.**40 to **9.**41
Crossrange resolution, **17.**1
Cross section. *See* radar cross section
Crosstalk, in tracking, **9.**40 to **9.**41
Crowbar, in transmitters, **10.**24
CryoSat altimeter, **18.**42 to **18.**43

D
D region, **20.**14
Dällenbach layer, **14.**37 to **14.**38
Data link, missile, **5.**26
Data links, in MFAR, **5.**24 to **5.**27
Data processing, **1.**3
dBZ, **19.**5 to **19.**6
DC operation of CFA, **10.**17, **10.**24
Deceptive ECM (DECM), and ECCM, **24.**6 to **24.**7, **24.**40
  and tracking radar, **24.**41 to **24.**42
Decimation, **6.**41 to **6.**42
Decimation filters, **25.**28 to **25.**32
Decoys, **24.**6, **24.**8
Delta-Sigma converters, **6.**36
Depressed collector, **10.**10
Detection acceptance, **7.**25 to **7.**26
Detectors
  batch processor, **7.**8 to **7.**11
  binary integrator, **7.**7 to **7.**8, **7.**12 to **7.**13
  M-out-of-N, **7.**7 to **7.**8
  moving window, **7.**4 to **7.**7
  nonparametric, **7.**17 to **7.**18
  optimal, **7.**2 to **7.**4
  practical, **7.**4 to **7.**11
  rank, **7.**17 to **7.**18
Dicke fix, **24.**33, **24.**35
Differential reflectivity, **19.**18
Diffraction, **26.**5
Digital beamforming, **13.**8, **13.**56 to **13.**57
  multiple beams, **25.**17 to **25.**19

Digital downconversion (DDC), **6.**41 to **6.**42, **6.**44 to **6.**45, **25.**6 to **25.**15
Digital filters, **25.**26 to **25.**32
Digital pulse compression, **25.**19
  generation of, **8.**28 to **8.**30
Digital receiver, **6.**40 to **6.**46
  block diagram, **25.**2
  direct sampling, **25.**38
Digital RF memory (DREM), **24.**6 to **24.**7
Digital signal processing
  hardware implementation, **25.**35 to **25.**37
  timing dependence, **25.**34 to **25.**35
  tools, **25.**22 to **25.**34
Digital upconverter (DUC), **25.**21 to **25.**22
Diode phase shifter, **13.**51 to **13.**52
Dipole antenna, **21.**25 to **21.**26
Direct digital downconversion, **25.**10 to **25.**15
Direct digital synthesizer (DDS), **6.**22, **6.**48 to **6.**49, **25.**20 to **25.**21
Direct sampling digital receiver. **25.**38
Direction finding (DF) and radar, **7.**50 to **7.**54
Directive gain, of an antenna, **12.**5
Discrete Fourier transform (DFT), **25.**32 to **25.**34
Displaced Phase Center Antenna (DPCA), **3.**10 to **3.**13, **3.**19 to **3.**21, **5.**2 to **5.**3
Distortion, in receiver, **6.**6 to **6.**7
Doppler ambiguity resolution, **4.**33 to **4.**34

Doppler beam sharpening (DBS), **5.**34 to **5.**36, **5.**37, **17.**2, **17.**3
Doppler, in bistatic radar, **23.**14 to **23.**17
Doppler filter bank, **2.**7 to **2.**9
Doppler filter straddle loss, **4.**43
Doppler filter weighting loss, **4.**43
Doppler radars, **4.**1
Doppler scintillation, in tracking radars, **9.**36 to **9.**37
Doppler shift, **1.**10
Doppler spectra, spurious, **6.**11
Doppler spectrum
  of ground echo, **16.**16 to **16.**19
  of HF sea echo, **20.**49 to **20.**52, **20.**75
Doppler weather radar, **1.**7
Ducts
  atmospheric, **26.**7 to **26.**10
  elevated, **26.**12 to **26.**13
  evaporation, **26.**11 to **26.**12
  surface, **26.**10 to **26.**11
Duplexer, **1.**2, **6.**2
Dynamic range, **1.**3, **6.**4 to **6.**8
  of A/D converters, **6.**38 to **6.**39
  and ECCM, **24.**32 to **24.**33
  in pulse doppler radar, **24.**32 to **24.**33

E
E-2C, **3.**3
E-2D, **3.**1 to **3.**3
E region, **20.**14
Early-late gate range tracking, **9.**21 to **9.**22
Eclipsing loss, **4.**40 to **4.**43, **5.**18
Effective-earth-radius model, **26.**15
Eldora radar, **19.**37 to **19.**38
EM system assessment, **26.**18 to **16.**23
Electronic Attack, **24.**2
Electronic counter-countermeasures (ECCM)
  antenna-related, **24.**10 to **24.**31
  decoys, **24.**8
  efficiency, **24.**54 to **24.**56
  and imaging radar, **24.**48 to **24.**52
  and ISAR, **24.**51 to **24.**52
  operational-deployment techniques, **24.**36 to **24.**37
  and over-the-horizon radar, **24.**52 to **24.**53
  and phased array radar, **24.**43 to **24.**48
  and radar equation, **24.**55 to **24.**56
  receiver related, **24.**32 to **24.**33
  role of the operator with, **24.**36
  SLB and SLC, **24.**17 to **24.**20
  sidelobe canceler, **24.**14 to **24.**20
  signal-processing related, **24.**33 to **24.**36
  STAP, **24.**21 to **24.**22
  and subarrays, **24.**25 to **24.**29
  and surveillance radars, **24.**37 to **24.**40
  and SAR, **24.**48 to **24.**51
  techniques, **24.**8 to **24.**9, **24.**37 to **24.**53
  terminology, **24.**2
  and tracking radars, **24.**40 to **24.**43
  transmitter related, **24.**31 to **24.**32
Electronic countermeasures (ECM), **24.**5 to **24.**8
Electronic Protection (EP), **24.**2
Electronic-scan monopulse, **9.**12 to **9.**13
Electronic Support (ES), **24.**2
Electronic Warfare Support Measures (ESM), **24.**2 to **24.**5
Elevated ducts, **26.**12 to **26.**13

Elint, **24.**2, **24.**3, **24.**4
Emission Control (EMCON), **24.**37
ENVIS AT, **18.**11
Errors in tracking radar
  external causes of, **9.**37 to **9.**43
  internal sources of, **9.**42 to **9.**43, **9.**47
  reduction techniques, **9.**46 to **9.**47
  sources of, **9.**26
  target caused, **9.**26 to **9.**37
ERS SAR, **18.**8 to **18.**9
Evaporation ducts, **26.**11 to **26.**12
Exponential model of land clutter, **2.**12 to **2.**16
Extended Interaction Klystron (EIK), **10.**11
External noise at HF, **20.**43 to **20.**45

F
F-117 low cross section aircraft, **14.**40 to **14.**42
F region, **20.**14 to **20.**15
Fading, **16.**12
False-alarm control, **7.**11 to **7.**19
Far field, **14.**4
Far-field criterion, **14.**27 to **14.**28
Fast Fourier transform (FFT), **25.**33 to **25.**34
  filter bank, **2.**55 to **2.**56
Fast Time Constant (FTC), in CMR, **22.**5
Fast-wave tube, **10.**3
Ferrite phase shifters, **13.**52 to **13.**53
Fighter aircraft missions, **5.**10 to **5.**16

Filter-bank design, **2.**52 to **2.**59
Filter mismatch loss, in MTI, **2.**22 to **2.**23
Filters, in radar, **6.**24 to **6.**29
Finite impulse response (FIR) filters, **2.**33, **25.**26 to **25.**28
FM-CW radar, **1.**5, **9.**20
Focused SAR, **17.**2
Foliage-penetration (FOPEN) SAR, **17.**33 to **17.**34
Fog, attenuation in, **19.**12
Forward scatter, in bistatic radar, **23.**21
Fourier transform, **8.**38
Frank codes, **8.**19 to **8.**20
Free-space propagation, **26.**13 to **26.**15
Frequency agility, **1.**9, **24.**31
Frequency diversity, **1.**9, **24.**31
Frequency, effect on radar, **1.**14 to **1.**18
Frequency multipliers, **6.**49 to **6.**50
Frequency synthesis, **6.**21 to **6.**22
Front end, of receiver, **6.**10 to **6.**14

G
GaAs PHEMT FET, **11.**12 to **11.**16
Geometrical optics, **14.**20 to **14.**21
  and ground echo, **16.**8
  reflector analysis, **12.**33
Geometric theory of diffraction, **14.**24 to **14.**25
Geosat, **18.**41 to **18.**42
Ghosts, **2.**88, **4.**33, **14.**35
Glint
  in bistatic radar, **23.**20 to **23.**21
  in tracking radar, **9.**30 to **9.**35
Global Nearest Neighbor (GNN), **7.**39 to **7.**40
Graceful degradation, **13.**4
Graded absorber, **14.**37
Grating lobes, **13.**10 to **13.**12
Gregorian antenna, **12.**21, **12.**25
Gregorian system, **14.**33 to **14.**34
Grid-controlled vacuum tubes, **10.**21 to **10.**23, **10.**25
Grid locking, **7.**49
Ground echo
  available information from, **16.**4
  bistatic, **16.**29
  fading of, **16.**12 to **16.**19
  at HF, **20.**29 to **20.**30
  imaging radar interpretation, **16.**55 to **16.**56
  at low grazing angle, **16.**52 to **16.**55
  measurement techniques for, **16.**19 to **16.**29
  in meteorological radar, **19.**13 to **19.**14
  models for scattering coefficient, **16.**29 to **16.**34
  near-grazing angle, **16.**52 to **16.**55
  near-vertical, **16.**25
  number of independent samples, **16.**24
  parameters affecting, **16.**4 to **16.**7
  polarimetry, **16.**46 to **16.**52
  scattering coefficient data, **16.**35 to **16.**46
  scattering coefficients from radar images, **16.**28 to **16.**29
  sea ice, **16.**44 to **16.**46
  snow, **16.**42 to **16.**44
  soil moisture, **16.**40 to **16.**41
  speckle, **16.**55
  theoretical models, **16.**7 to **16.**12
  vegetation, **16.**41
Ground moving target indication (GMTI), **5.**38 to **5.**42, **17.**25
Ground moving target thresholding, **5.**39 to **5.**40

Ground moving target track (GMTT), **5.**38
Ground moving target weapon delivery, **5.**40 to **5.**41
Ground penetrating radar (GPR)
  antennas, **21.**24 to **21.**30
  applications of, **21.**3, **21.**35 to **21.**40
  archeological applications, **21.**36
  and land mines, **21.**37
  AT mines, **21.**34, **21.**39
  attenuation with, **21.**7 to **21.**9
  bandwidth, **21.**1
  characteristics of, **21.**4
  clutter in, **21.**5, **21.**10
  comparison with optical image, **21.**4
  depth resolution, **21.**11
  description of, **21.**1 to **21.**6
  dispersion in, **21.**11
  earth material properties, **21.**18 to **21.**19
  example of, **21.**2
  forensic investigations, **21.**35
  and frequency domain, **21.**23
  and glacier, **21.**37
  image processing, **21.**30 to **21.**35
  licensing of, **21.**39 to **21.**40
  modeling of, **21.**13 to **21.**18
  modulation techniques, **21.**21 to **21.**24

polarization, **21.**10 to **21.**11
propagation, **21.**6 to **21.**13
reflections in, **21.**9 to **21.**10
resolution, **21.**13
road thickness measurement, **21.**37 to **21.**38
and SAR, **21.**13
signal processing, **21.**30 to **21.**35
soil suitability map, **21.**20
from space, **18.**59 to **18.**62
systems, **21.**20 to **21.**21
velocity of propagation, **21.**11
Ground plane, **14.**31
Guard blanking loss, **4.**44
Guard channel, in pulse doppler, **4.**19 to **4.**22
Gyrotrons, **10.**3, **10.**17 to **10.**19, **10.**26

H
Hail, attenuation in, **19.**11 to **19.**12
Hail, detection of, **19.**31
Height measurement, with InSAR, **17.**30 to **17.**33
Height measurement, with SAR, **17.**27 to **17.**33
HF, **1.**15
HF over-the-horizon radar. *See* over-the-horizon radar
High-medium PRF, **4.**7
High-PRF pulse doppler, **4.**7 to **4.**8
High-PRF range-while-search, **4.**36
High-PRF ranging, in pulse doppler, **4.**34 to **4.**35
High-resolution radar, **1.**5
Hilbert transform, **6.**43 to **6.**44
HJ-1-C SAR, **18.**13
Hot clutter, **24.**43
Hybrid models for propagation, **26.**16 to **26.**17
Hybrid processors, **25.**37

I
Identification friend or foe (IFF), integration with radar, **7.**50
I/Q channels, **6.**31 to **6.**35
Image-reject mixer, **6.**13
Imaging radar ground echo, **16.**55 to **16.**56
Improvement factor, for AMTI, **3.**8
Improvement factor limitations caused by staggering, **2.**42 to **2.**44
Improvement factor, for MTI, **2.**19 to **2.**20, **2.**23 to **2.**25, **6.**17 to **6.**18
Inductive output tube, **10.**22
Infinite impulse response (IIR) filters, **2.**33, **25.**26 to **25.**27
Insets, radar cross section of, **14.**11 to **14.**12
Instabilities in MTI, **2.**65, **2.**73
Instability limitations in MTI, **2.**72
Instantaneous bandwidth, **6.**9
Interclutter visibility in MTI, **2.**22
Interacting multiple model (IMM) in ADT, **7.**35 to **7.**37
Interferometric SAR (InSAR), **17.**5, **17.**23 to **17.**24
target height measurement, **17.**30 to **17.**33
International Maritime Organization (IMO), **22.**1, **22.**4, **22.**8, **22.**27, **22.**28, **22.**32
Interpolation filters, **25.**28
Inverse Cassegrain, **9.**25
Inverse SAR (ISAR), **5.**23 to **5.**24, **5.**31 to **5.**33, **9.**37, **17.**5
and ECCM, **24.**51 to **24.**52
Ionogram, **20.**16 to **20.**17
Ionosphere, **20.**13 to **20.**21
Ionospheric models, **20.**19 to **20.**20
Isodoppler contours in bistatic radar, **23.**16

J
Jamming, **5.**27, **24.**5
Jaumann absorber, **14.**36 to **14.**37
J-ERS SAR, **18.**9
Jet-engine modulation, **5.**23
JianBing-5 SAR, **18.**12
Jindalee OTH radar, **20.**3, **20.**22
  performance model, **20.**67 to **20.**70
Joint STARS, **13.**66 to **13.**67, **17.**23 to **17.**24
JORN OTH, **20.**12, **20.**19

K
K band, **1.**17
Kalman filter, **7.**28 to **7.**35
Klystrode, **10.**22
Klystron, **10.**5 to **10.**8, **10.**26
  origin of, **10.**2 to **10.**3

L
L band, **1.**16
Land clutter. *See* ground echo
Laser radar, **1.**18

Letter-band nomenclature, **1.**13 to **1.**14
Lewis and Kretschmer codes, **8.**20 to **8.**22
Limiters, **6.**29 to **6.**31
Line-type modulator, **10.**23
Linear array antenna, **13.**11 to **13.**13
Linear-beam vacuum tube, **10.**3
Linear-beam amplifier, **10.**4 to **10.**13
Linear FM pulse compression, **8.**3 to **8.**11
Liquid crystal display, for CMR, **22.**19
Lobing, in elevation, **22.**6 to **22.**7
Local oscillator, **6.**14 to **6.**22
Low-earth orbit, **18.**2
Low grazing-angle clutter measurements, **16.**52 to **16.**55
Low-noise amplifier, **6.**10
Low radar cross section
  ships **14.**42 to **14.**43
  vehicles, **14.**39 to **14.**43
Low-sidelobe antenna, **24.**10 to **14.**11

M
Madre HF OTH radar, **20.**24 to **20.**45
Magellan Venus mapper, **18.**48 to **18.**52
Magnetron, **10.**2, **10.**14 to **10.**16, **10.**26
  for civil marine radar, **10.**16, **22.**12
Main-beam cancellation, **24.**33
MAPSAR, **18.**14 to **18.**15
Marcum Q-function, **24.**22
Marine boundary layer, **26.**12
Matched filter, **1.**3, **6.**25, **8.**39 to **8.**40
  loss in pulse doppler, **4.**40
Maximal-length sequences, **8.**18
Maxwell's equations, **14.**17
Mechanical design of antenna, **12.**35 to **12.**41
Medium PRF for air-to-air mode, **5.**16 to **5.**20
Medium PRF for pulse doppler, **4.**8
Medium PRF for range while search, **4.**37 to **4.**38
Medium PRF selection algorithms, **5.**18 to **5.**20
Metal plate, radar cross section of, **14.**8 to **14.**9
Meteor trails, and HF OTH radar, **20.**38 to **20.**40
Meteorological radar
  airborne, **19.**37 to **19.**38
  attenuation effects, **19.**7 to **19.**12
  calibration, **19.**18 to **19.**19
  design considerations, **19.**6 to **19.**19
  ground clutter effects, **19.**13 to **19.**14
  measurement accuracy, **19.**22 to **19.**23
  microbursts, **19.**29 to **19.**31
  MTI filter design for, **2.**46 to **2.**51
  multiple radars, **19.**33 to **19.**35
  operational applications, **19.**25 to **19.**33
  phased array, **19.**35 to **19.**37
  polarization, **19.**18, **19.**27 to **19.**28, **19.**33
  precipitation measurement, **19.**26 to **19.**28
  processor implementation. **19.**24 to **19.**25
  pulse compression, **19.**23 to **19.**24
  pulse-pair algorithm, **19.**21
  range and velocity ambiguities, **19.**12 to **19.**13
  research applications, **19.**33 to **19.**40
  severe storm warning, **19.**28 to **19.**33
  signal processing, **19.**19 to **19.**25
  spaceborne, **19.**38 to **19.**39
  spaced antenna techniques, **19.**40
  spectrum moment estimation, **19.**20 to **19.**21

typical designs, **19.**14 to **19.**18
whitening filter, **19.**24
wind profiler, **19.**39 to **19.**40
Method of moments, **14.**1, **14.**18 to **14.**19
MFAR. *See* Multifunctional Fighter Aircraft Radar
Microbursts, **19.**29 to **19.**31
Microwaves, **1.**1, **1.**14
Microwave monolithic integrated circuits (MMIC)
characteristics of, **11.**24 to **11.**29
power amplifiers, **11.**26
low-noise amplifiers, **11.**26
phase shifters, **11.**27 to **11.**29
transmit/receive switching, **11.**27
Microwave power module (MPM), **10.**13
Military radar, **1.**20
Millimeter waves, **1.**14, **1.**17 to **1.**18
Mirror-scanned antenna, **9.**25
Missile guidance performance assessment, **5.**42
Missile-range instrumentation radar, **9.**2
Mixer, image-reject, **6.**13
Mixers, performance parameters, **6.**14
Mixers, spurious response of, **6.**11 to **6.**13
MMIC. *See* Microwave monolithic integrated circuits

Modeling, of propagation, **26.**13 to **26.**17
Modified generalized sign test processor, **7.**17 to **7.**19
Modulators, **10.**23 to **10.**25
Monopulse-antenna feed techniques, **9.**6 to **9.**10
Monopulse antenna feeds, **12.**26 to **12.**28
Monopulse, dual band, **9.**24 to **9.**25
Monopulse, and ECM, **24.**23 to **24.**24, **24.**42 to **24.**43
Monopulse, phased arrays, **13.**4
Monopulse tracking, **9.**3 to **9.**16
Moving Target Indication (MTI), **1.**5, **4.**2
  A/D converter effect on dynamic range, **2.**78 to **2.**80
  adaptive, **2.**80 to **2.**83
  binomial weight cancelers, **2.**35
  and birds as clutter, **2.**85, **2.**87 to **2.**88, **2.**96 to **2.**98
  blind speeds, **2.**9 to **2.**10
  block diagram, **2.**4 to **2.**7
  canonical filter design, **2.**34
  clutter attenuation (CA), **2.**20 to **2.**21
  clutter characteristics, **2.**10 to **2.**19
  clutter filter-bank design, **2.**52 to **2.**59
  clutter filter response, **2.**9 to **2.**10
  clutter maps, **2.**83 to **2.**87
  clutter visibility factor (V_{oc}), **2.**23
  definitions, **2.**19 to **2.**23
  dynamic range, **2.**78 to **2.**80
  environmental considerations, **2.**94 to **2.**100
  feedforward canceler, **2.**39
  filter design, **2.**25 to **2.**46
  filter design for weather radar, **2.**46 to **2.**51
  filter mismatch loss, **2.**22 to **2.**23
  hardware considerations, **2.**92 to **2.**94
  improvement factor, **2.**19 to **2.**20, **2.**23 to **2.**25, **6.**17 to **6.**18
  interclutter visibility (ICV), **2.**22
  limitation due to scanning, **2.**23, **2.**38
  limiting in receiver, **2.**59 to **2.**65
  one-delay canceler, **2.**35
  optimum clutter filter design, **2.**25 to **2.**33
  performance degradation due to limiting, **2.**59 to **2.**65
  pulse compression considerations, **2.**75 to **2.**78
  purpose of, **2.**2
  radial velocity ambiguity resolution, **2.**89 to **2.**91
  range ambiguity resolution, **2.**89 to **2.**91
  rules for, **2.**92 to **2.**94
  in SAR, **17.**23
  and STC, **2.**96 to **2.**98
  sensitivity velocity control (SVC), **2.**87 to **2.**91, **2.**99 to **2.**100
  signal-to-clutter ratio improvement (I_{SCR}), **2.**21 to **2.**22
  stability requirements, **2.**65 to **2.**78
  staggered PRF design, **2.**39 to **2.**46
  subclutter visibility (SCV), **2.**6, **2.**22
  superclutter visibility, **2.**84
  three-delay canceler, **2.**37
  two-delay canceler, **2.**36
  unwanted targets, removal of, **2.**96 to **2.**100
Moving targets, in SAR, **17.**23 to **17.**27
Moving Target Detection (MTD), **2.**6 to **2.**9
  block diagram, **2.**6 to **2.**7
  ducted propagation, effect of, **2.**95
Moving-window detector, **7.**4 to **7.**7
MTBF, of klystrons, **10.**6 to **10.**7
Multi-beam digital beamforming, **25.**17 to **25.**19
Multifunction radar, **1.**7, **13.**1
Multifunctional Fighter Aircraft Radar (MFAR), **5.**2 to **5.**7
  processing, **5.**5
  range-doppler situation in, **5.**7 to **5.**8

software structure, **5.**6 to **5.**7
Multipath, **26.**4 to **26.**5
in tracking radar, **9.**37 to **9.**40, **9.**46 to **9.**47
Multiple-beam klystron, **10.**7 to **10.**8
Multiple hypothesis algorithm in track association, **7.**40 to **7.**41
Multiple Object Tracking Radar (MOTR), **9.**2, **9.**26
Multiple reflector antenna, **12.**21 to **12.**24
Multiple scattering in target recognition, **14.**35
Multiplicative noise, **4.**28, **18.**18
in SAR, **17.**17 to **17.**18

N

NCAR dual band radar, **19.**17
Networked radar, **7.**46 to **7.**49

Nexrad (WSR-88D) weather radar, **19.**1, **19.**16 to **19.**17
  automated weather products, **19.**25
Noise, in receiver, **6.**4 to **6.**5
Noise, in tracking, **9.**27 to **9.**30
Noncooperative air target recognition, **5.**22 to **5.**24
Nonlinear FM pulse compression, **8.**12 to **8.**16
Nonparametric detectors, **7.**17 to **7.**18
Nonspecular absorbers, **14.**38
*n*th-time-around tracking, **9.**24
Nyquist, **25.**2, **25.**4 to **25.**6

O
Ogive, radar cross section of, **14.**7 to **14.**8
On-axis tracking **9.**25 to **9.**26
Operating bandwidth, **6.**9
Optics region, **14.**5 to **14.**6
Optimal detection, **7.**2 to **7.**4
Orographic rain, **19.**26
Oscillator vs. amplifier transmitter, **10.**4
Ovals of Cassini, **23.**6 to **23.**8
Over-the-horizon radar
  antennas for, **20.**24 to **20.**26, **20.**45 to **20.**46
  calibration, **20.**48 to **20.**49
  clutter, **20.**29 to **20.**33
  coherent processing for, **20.**6
  compared to microwave radar, **20.**7 to **20.**10
  doppler spectrum of the sea, **20.**49 to **20.**52
  and ECCM, **24.**52 to **24.**53
  examples of, **20.**11 to **20.**12
  external noise in, **20.**40
  and meteor trails, **20.**38 to **20.**40
  noise models, **20.**43 to **20.**45
  oceanography, **20.**33 to **20.**38
  performance modeling, **20.**55 to **20.**70
  propagation factor, **20.**6 to **20.**7
  radar cross section, **20.**26 to **20.**29
  radar equation, **20.**5 to **20.**7
  receiving system, **20.**45 to **20.**49
  resource management, **20.**54 to **20.**55
  revisit times, **20.**4 to **20.**5
  signal processing, **20.**49 to **20.**54
  sky wave radar design, **20.**8 to **20.**13
  spectrum occupancy, **20.**40 to **20.**45
  surface-wave radar, **20.**70 to **20.**76
  tracking, **20.**53 to **20.**54
  transmitters, **20.**23 to **20.**26
  waveforms, **20.**21 to **20.**23

P
PALSAR, **18.**11, **18.**15 to **18.**16
Parabolic cylinder antenna, **12.**18 to **12.**19
Parabolic equation methods, for propagation, **26.**16
Parabolic reflector antenna, **12.**17 to **12.**18
Passive bistatic radar, **23.**29
Passive ECM, **24.**5 to **24.**6
Passive listening, **5.**34
Patriot radar, **13.**66 to **13.**67
Pave Paws radar, **11.**31 to **11.**32
Penetration aids accompanying ballistic missiles, **24.**6
Phase-comparison monopulse, **9.**11 to **9.**12
Phase instabilities, **2.**66 to **2.**70

Phase noise in pulse doppler, **4.**28 to **4.**30
Phase shifters, **13.**51 to **13.**53
Phased array radar
  active aperture, **13.**53 to **13.**55
  antennas, **13.**2 to **13.**3
  AN/SPY-1, **13.**62
  AN/SPY-3, **13.**69
  aperture matching, **13.**20
  array simulator for, **13.**25 to **13.**26
  bandwidth of, **13.**38 to **13.**45
  beam switching, **13.**8
  calibration of active arrays, **13.**60 to **13.**62
  circular polarization, **13.**6
  constrained feed, **13.**46
  Counter Battery Radar (COBRA), **13.**62
  digital beamforming, **13.**56 to **13.**57
  diode phase shifters, **13.**51 to **13.**52
  and ECM, **24.**43 to **24.**48
  element pattern, **13.**22 to **13.**23
  errors in, **13.**30 to **13.**38
  feed networks, **13.**46 to **13.**50
  ferrite phase shifters, **13.**52 to **13.**53
  frequency scan, **13.**7 to **13.**8
  gain, **13.**13 to **13.**15
  grating lobes, **13.**10 to **13.**12, **13.**17 to **13.**19
  ground-based, **13.**63 to **13.**65
  illumination functions, **13.**28 to **13.**29
  instantaneous bandwidth, **13.**42 to **13.**45
  limited scan, **13.**6 to **13.**7

linear array, **13.**11 to **13.**13
low sidelobes, **13.**28 to **13.**33
meteorological, **19.**35 to **19.**37
monitoring of, **13.**4 to **13.**5
monopulse tracking, **9.**12 to **9.**13, **13.**4
multifunction, **13.**1
mutual coupling, **13.**20 to **13.**22
optical feed, **13.**46
parallel feed, **13.**48 to **13.**49
periodic errors in, **13.**35 to **13.**38
phase-only control, **13.**58
phase quantization, **13.**34
phase shifters, **13.**51 to **13.**53
planar array, **13.**15 to **13.**19
radiation pattern nulling, **13.**57 to **13.**60
scanning, **13.**7 to **13.**9
series feed, **13.**47 to **13.**48
simultaneous receive beams, **13.**54 to **13.**56
small arrays, **13.**27
solid-state modules for, **13.**53 to **13.**54
solid-state transmitter for, **11.**24 to **11.**31
subarrays, **13.**43 to **13.**44, **13.**49 to **13.**50
surface waves and mutual coupling, **13.**24 to **13.**45
Taylor illumination, **13.**29
thinned arrays, **13.**23 to **13.**24
3D search, **13.**4
theory, **13.**9 to **13.**15
time-delay networks, **13.**44 to **13.**45
time-delay scanning, **13.**7
tracking with, **7.**46
Volume Search Radar (VSR), **13.**62 to **13.**63
wide bandwidth operation, **13.**6
Phase shift in DSP, **25.**22 to **25.**25
Physical optics, **14.**21 to **14.**24
and ground echo, **16.**9 to **16.**10
reflector analysis, **12.**31 to **12.**33
Physical theory of diffraction, **14.**25 to **14.**26
Pilotage and CMR, **27.**31
Pioneer Venus, **18.**44
Planar array, **13.**15 to **13.**19
Planetary radars, **18.**43 to **18.**53
atmospheric sounding, **18.**62 to **18.**63
Cassini, **18.**46
Clementine, **18.**46 to **18.**47
cloud profiling, **18.**63
flight systems, **18.**43, **18.**56 to **18.**58
ice exploration, **18.**47 to **18.**48
ionospheric sounding, **18.**62
Magellan, **18.**45 to **18.**46, **18.**48 to **18.**52
polarization, **18.**52
rainfall measurement, **18.**62
scatterometers, **18.**53 to **18.**58
sounders, **18.**59 to **18.**63
table of, **18.**44
Venera, **18.**43 to **18.**45
Plasma frequency, **20.**2
Platform motion, in AMTI, **3.**3 to **3.**14
compensation abeam, **3.**10 to **3.**14
forward direction, **3.**21 to **3.**23
and scan compensation, **3.**18 to **3.**21
P(n,k) polyphase codes, **8.**22 to **8.**24
Point-clutter scatterers, **2.**18 to **2.**19
Polarimetric SAR, **17.**22
Polarization
and ground echo, **16.**46 to **16.**52

in ground penetrating radar, **21.**10 to **21.**11
in meteorological radar, **19.**18
for planetary radar, **18.**52
Polarization-twist reflector, **12.**23
Polyphase codes, **8.**19 to **8.**24
Power-aperture product, **10.**1
Power supply, for AESA, **5.**10
Practical detectors, **7.**4 to **7.**11
Precipitation measurement, **19.**26 to **19.**28
Precision velocity update, **5.**33 to **5.**34
Probabilistic data association, **7.**39 to **7.**40
Probability of detection, in pulse doppler, **4.**46 to **4.**48
Probability of false alarm, in pulse doppler, **4.**44 to **4.**46
Propagation, anomalous, **26.**6 to **26.**13
Propagation factor, **26.**1
Propagation, in free space, **26.**13 to **26.**15
Propagation loss, **4.**40
Propagation modeling, **26.**13 to **26.**17
Propagation, standard, **26.**4 to **26.**6
Pseudo coherent radar, **6.**20
Pulse chasing in bistatic radar, **23.**28 to **23.**29
Pulse compression
block diagram, **8.**1 to **8.**2
comparison of various waveforms, **8.**27
digital, **25.**19
and ECCM, **24.**34 to **24.**35
examples of, **8.**30 to **8.**36
factors affecting choice of, **8.**26 to **8.**27

Pulse compression (*cont.*)
implementation of, **8.**28 to **8.**30
linear FM, **8.**3 to **8.**11
in meteorological radar, **19.**23 to **19.**24
and MTI, **2.**75 to **2.**78
nonlinear FM, **8.**12 to **8.**16
phase coded, **8.**16 to **8.**24
radar, **1.**5
range sidelobes, **6.**29
SAW devices for, **8.**10 to **8.**11
signal analysis definitions, **8.**36 to **8.**37
stretch, **8.**31 to **8.**36
time-frequency coded, **8.**25 to **8.**26
waveforms, **8.**2 to **8.**26
Pulse doppler
antenna, **4.**12 to **4.**13
applications, **4.**2 to **4.**3
basic configuration, **4.**10 to **4.**14
block diagram, **4.**11
clutter, **4.**14 to **4.**24
clutter-limited detection, **4.**48
doppler ambiguity resolution, **4.**33 to **4.**34
dynamic range, **4.**24 to **4.**27
high-PRF ranging, **4.**34 to **4.**35
losses in, **4.**39
nomenclature, **4.**1 to **4.**2
probability of detection, **4.**46 to **4.**48
probability of false alarm, **4.**44 to **4.**46
range ambiguity resolution, **4.**31 to **4.**33
range performance, **4.**39 to **4.**48
search mode, **4.**36 to **4.**38
spectrum, **4.**4 to **4.**6
stability requirements, **4.**27 to **4.**31
timeline definitions, **4.**9 to **4.**10
track mode, **4.**38 to **4.**39
Pulse-pair algorithm for meteorological radar, **19.**21
Pulse repetition frequency
and ambiguities in pulse doppler, **4.**6 to **4.**10
and doppler, **4.**2 to **4.**4
for SAR, **17.**13 to **17.**15
Pyramidal absorber, **14.**37 to **14.**38

Q
Quadraphase codes, **8.**24
Quantization noise, affect on MTI, **2.**73 to **2.**74

R
Racons, and CMR, **22.**26 to **22.**27
Radar
applications of, **1.**20 to **1.**22
bandwidth in, **1.**8 to **1.**9
basic parts of, **1.**2
beacons, in CMR, **22.**25
blinking for ECCM, **24.**37
block diagram, **1.**3
in brief, **1.**1
conceptual system design, **1.**22 to **1.**23
cross section
of aircraft, **14.**13 to **14.**14
approximate methods, **14.**19 to **14.**27
approximations for simple scatterers, **14.**10

basic echo mechanisms, **14.**2 to **14.**4
of birds, **14.**11
bistatic, **23.**19 to **23.**21
of bodies of revolution, **14.**39 to **14.**40
characteristics, **14.**5 to **14.**16
of complex objects, **14.**11 to **14.**16
of a corner reflector, **14.**9 to **14.**11
definition of, **14.**4 to **14.**5
exact methods, **14.**16 to **14.**19
at HF, **20.**26 to **20.**29
general, for various target types, **14.**16
of insects, **14.**11 to **14.**12
of a man, **14.**11 to **14.**12
measurement ranges, **14.**30 to **14.**35
measurement techniques, **14.**27 to **14.**35
of a metal plate, **14.**8 to **14.**9
of an ogive, **14.**7 to **14.**8
prediction techniques, **14.**16 to **14.**27
of ships, **14.**13 to **14.**15
of a short wire dipole, **14.**7
of a sphere, **14.**5 to **14.**6
of string, **14.**6, **14.**7, **14.**29
and surface traveling waves, **14.**7
of target support, **14.**29 to **14.**30
doppler shift in, **1.**10
echo suppression, **14.**36 to **14.**43
in A-12, **14.**40
in B-2, **14.**42
in F-117, **14.**40 to **14.**42
in SR-71, **14.**39 to **14.**41
in X-45C, **14.**42 to **14.**43
by absorbers, **14.**36 to **14.**38

by shaping, **14.**36, **14.**38 to **14.**39
in ships, **14.**42 to **14.**43
equation, **1.**10 to **1.**13
bistatic, **23.**4 to **23.**6
and chaff, **24.**55 to **24.**56
in conceptual design, **1.**23
ground penetrating radar, **21.**8 to **21.**9
HF over-the-horizon, **20.**5 to **20.**7
and jamming, **24.**55
for meteorological targets, **19.**3 to **19.**6
surveillance, **3.**1
frequency bands, **1.**13 to **1.**14
hole, in propagation, **26.**9
information from, **1.**7 to **1.**10
letter-band nomenclature, **1.**13 to **1.**14
multifunction, for fighter aircraft, **5.**1 to **5.**10
networked, **7.**46 to **7.**49
nomenclature, **1.**18 to **1.**19
oceanography at HF, **20.**30 to **20.**38
past advances in, **1.**19 to **1.**20
reflectivity, **19.**3
reflectivity factor, Z, **19.**4
scheduling and control, **7.**46
types of, **1.**5 to **1.**7
warning receiver (RWR), **24.**4 to **24.**5
work station, **24.**55 to **24.**56
RADARS AT, **18.**9 to **18.**10
Radial velocity, **1.**7
Radiating elements, **13.**5 to **13.**6
Radomes, **12.**39 to **12.**41
Radome loss, **4.**40
Rain, attenuation by, **19.**8 to **19.**11
Rain, detection of, **19.**3 to **19.**6
RAMP air traffic control radar, **11.**33 to **11.**34
Range, as used in radar, **1.**2, **1.**7
Range
ambiguities, in meteorological radar, **19.**12 to **19.**13
ambiguity resolution in pulse doppler, **4.**31 to **4.**33
and doppler, in MFAR, **5.**7 to **5.**8
error, in tracking radar, **9.**43 to **9.**44
filter map in sidelobe blanking, **24.**13 to **24.**14
-gate pull-off, **24.**6, **24.**44
-gate straddle loss, **4.**40 to **4.**43
-gated high PRF (RGHPRF), **5.**20 to **5.**22
gates, **4.**2
gating in pulse doppler, **4.**9
glint, **9.**35 to **9.**36, **9.**47
noise, **9.**35 to **9.**36
sidelobes, **6.**29
tracking, **9.**21 to **9.**24
Rank detector, **7.**17
Rapid Doppler on Wheels (Rapid-DOW) **19.**36 to **19.**37
Ratio detector, **7.**13 to **7.**14
Rayleigh region, **14.**5, **19.**3 to **19.**4
Rayleigh scattering, **14.**19 to **14.**20
Ray tracing, **20.**20 to **20.**21
Receiver, **1.**3
analog-to-digital converter, **6.**35 to **6.**40
bandwidth, **6.**9
channel matching, **6.**29
COHO, **6.**20
configuration of, **6.**1 to **6.**4
digital, **6.**40 to **6.**46
diplex operation, **6.**46 to **6.**47
dynamic range, **6.**4 to **6.**8

effect on radiated signal, **6.**11
exciter, **6.**47
filtering, **6.**24 to **6.**29
front end, **6.**10 to **6.**14
gain control of, **6.**22 to **6.**24
instability, **6.**20 to **6.**21
I/Q channels, **6.**31 to **6.**35
limiter, **6.**29 to **6.**31
local oscillator, **6.**14 to **6.**22
multi-channel, **6.**45 to **6.**46
noise, **6.**4 to **6.**5
and tracking accuracy, **9.**42
spurious response, **6.**11 to **6.**12
STALO, **6.**14 to **6.**20
upconversion, **6.**47 to **6.**50
waveform generation, **6.**47 to **6.**50
Receiver-related ECCM, **24.**32 to **24.**33
Reentrant structures and cross section, **14.**3
Reflectivity, **19.**3
Reflectivity factor, Z, **19.**4
Reflector antenna
analysis, **12.**31 to **12.**35
applications of, **12.**1 to **12.**2
architecture, **12.**16 to **12.**25
array feeds for, **12.**28 to **12.**30
basic principles of, **12.**3 to **12.**15
environmental factors, **12.**39
feed blockage, **12.**6 to **12.**8

Reflector antenna (*cont.*)
feed displacement, **12.**14
feeds for, **12.**25 to **12.**30
gain optimization, **12.**8 to **12.**10
mechanical design, **12.**35 to **12.**41
radomes for, **12.**39 to **12.**41
role in radar, **12.**1
spillover loss, **12.**6
strut blockage, **12.**14 to **12.**15
surface ace of, **12.**10 to **12.**14
surface roughness loss, **12.**12 to **12.**14
types of, **12.**2
Refraction, **14.**20, **26.**3 to **26.**4
Refractivity, **26.**3 to **26.**4
Refractivity measurements, **19.**32 to **19.**33
Region of uncertainty (ROU) in target tracking, **7.**47 to **7.**48
Remote sensing of the environment, **1.**21
Remote sensing, with SBR, **18.**11
Repeater jammer, **24.**6
Resolution, in GPR, **21.**11 to **21.**13
Resonance region, **14.**5
RF receive loss, **4.**40
RF transmit loss, **4.**40
RGHPRF (Range Gated High PRF) algorithm, **5.**22
Ricker wavelet, **21.**11, **21.**21
RIS AT SAR, **18.**14
River radar, **22.**9
ROTHR, **20.**3

S
S-193 altimeter, **18.**33
S band, **1.**16
Salisbury screen, **14.**36
Sampled signal spectrum, **25.**3 to **25.**5
Sampling period in phased arrays, **24.**45
Sampling receiver, **21.**23
SAR-Lupe radar, **18.**13 to **18.**14
SARTs (Search and Rescue Transponders), **22.**27
SAW delay line for pulse compression, **8.**10 to **8.**11
Scan compensation, and platform motion, **3.**18 to **3.**21
Scan SAR, **18.**12 to **18.**13, **18.**24
Scatterer, **14.**4
Scatterometers, **16.**19 to **16.**24, **16.**26 to **16.**28
space based, **18.**53 to **18.**58
table of, **18.**56
wind measurement, **18.**54 to **18.**55
Sea clutter
Bragg scatter, **15.**28 to **15.**32, **15.**38
breaking waves, **15.**6 to **15.**7
composite-surf ace model, **15.**30 to **15.**32
contaminants, effect of, **15.**26 to **15.**27
and ducting, **15.**24 to **15.**25
empirical behavior of, **15.**7 to **15.**27
as a global boundary-value problem, **15.**27 to **15.**32
at HF, **15.**19, **20.**30 to **20.**33
at high grazing angles, **15.**16
at low grazing angles, **15.**16 to **15.**18
at millimeter waves, **15.**20
numerical methods for, **15.**36
rain, effect of, **15.**23 to **15.**24
sea spikes, **15.**2, **15.**16 to **15.**17, **15.**35
shadowing, effect of, **15.**25
sigma zero definition, **15.**7

spectrum, **15.**20 to **15.**23
statistics, **15.**8 to **15.**9
surface currents, effect of, **15.**25 to **15.**26
surface features, **15.**33 to **15.**34
theories of, **15.**27 to **15.**37
wind speed and direction, **15.**12 to **15.**16
Sea descriptors, general, **15.**5 to **15.**6
Sea echo. *See* sea clutter
Sea ice, **16.**44 to **16.**46, **16.**48, **22.**14
Sea spikes, **15.**2, **15.**16 to **15.**17, **15.**35
Sea state, **15.**6, **22.**14
Sea surface, **15.**3
Sea surface search, **5.**30 to **5.**31
Seawinds scatterometer, **18.**58
Sensitivity time control (STC), **2.**96 to **2.**98, **6.**22 to **6.**24, **19.**6
in CMR, **22.**5, **22.**13
and cosecant-squared antenna, **2.**98
in pulse doppler, **4.**22
Sensitivity velocity control (SVC), **2.**87 to **2.**91, **2.**99 to **2.**100
Sensor integration, other than radar, **7.**49 to **7.**54
Senrad broadband radar, **24.**31
Sequential lobing, **9.**16 to **9.**17
Servosystems, for tracking radar, **9.**17 to **9.**19
Shaped pulses for transmitter, **10.**19 to **10.**20
Shaped reflector antennas, **12.**19 to **12.**20

Shaping, for radar cross-section reduction, **14.**38 to **14.**39
Ship, radar cross section, **14.**13 to **14.**15
Ships, low cross section, **14.**42 to **14.**43
Sidelooking airborne radar (SLAR), **1.**6 **17.**3
Sidelobe blanking (SLB), **24.**11 to **24.**14
Sidelobe canceler, **24.**14 to **24.**20
Sigma zero (for clutter echo), **15.**7, **16.**1 to **16.**3
Signal-to-clutter ratio improvement, **2.**21 to **2.**22
Signal-to-noise ratio, **1.**9
  in SAR, **17.**16 to **17.**17
Signal processing, and ECCM, **24.**33 to **24.**36
Signal processing, for meteorological radar, **19.**19 to **19.**25
Signal processor, **1.**3
Signal sampling, **25.**15 to **25.**16
Silicon bipolar junction transistor, **11.**10 to **11.**11
Silicon LDMOS FET, **11.**11 to **11.**12
Silicon transistor capabilities, **11.**13
Simultaneous lobing, **9.**3
SIR-C, **16.**46
SIR (spaceborne) SARs, **18.**8
Skin depth, **21.**8
Skip zone, **20.**3
Sky wave OTH radar, **20.**8 to **20.**13
Slip-SAR, **18.**14
Slow-wave tubes, **10.**3
Sniff (or passive listening), **5.**34
Snow, echo from, **16.**42 to **16.**44
Software structure, in MFAR, **5.**6 to **5.**7
Soil, dielectric properties of, **21.**19
Soil moisture, effect on radar echo, **16.**38, **16.**40 to **16.**41
Solar calibration of radar, **19.**19
SOLAS (Safety of Life at Sea), **22.**8
Solid-state
  advantages of, **11.**1 to **11.**5
  amplifier, **10.**4, **10.**26 to **10.**28
  amplitude and phase sensitivities, **11.**22 to **11.**23
  AN/SPS-40 transmitter, **11.**32 to **11.**33
  class of (amplifier) operation, **11.**18 to **11.**20
  devices, **11.**5 to **11.**17
  examples of, **11.**31 to **11.**34
  MMIC, **11.**24 to **11.**29
  modulator, **11.**24 to **11.**25
  modules, **13.**53 to **13.**54
  Pave Paws radar, **11.**31 to **11.**32
  performance capability, **11.**2
  phased array transmitter, **11.**24 to **11.**31
  power combining, **11.**20 to **11.**22
  power limitations, **11.**8 to **11.**10
  RAMP radar, **11.**33 to **11.**34
  spectral emissions, **11.**23
  transmit/receive modules, **11.**29 to **11.**31
Sounders, **18.**59 to **18.**63
Space-based radars (SBR)
  altimeters, **18.**29 to **18.**43
    block diagram, **18.**40
    CryoSat, **18.**42 to **18.**43
    flight systems, **18.**33 to **18.**37
    Geosat, **18.**41 to **18.**42
    orbit considerations, **18.**37 to **18.**38
    overview, **18.**30 to **18.**31
    precision, **18.**31
    theoretical foundations, **18.**38 to **18.**41
  hardware, **18.**4 to **18.**5
  meteorological, **19.**38 to **19.**39
  planetary. *See* planetary radars

synthetic aperture radars (SAR), **18.**5 to **18.**29
ambiguities, **18.**17 to **18.**18
ambiguity limits, **18.**22 to **18.**24
antennas, **18.**19
applications, **18.**29
data products, **18.**21 to **18.**22
data rate, **18.**20
design issues, **18.**16 to **18.**24
interferometry, **18.**24 to **18.**27
Kosmos, **18.**8
list of, **18.**6 to **18.**7
multiple channels, **18.**24 to **18.**29
nadir, return from, **18.**18
orbit characteristics, **18.**2 to **18.**4
polarimetry, **18.**27 to **18.**29
PRF constraints, **18.**16 to **18.**17
processing, **18.**20 to **18.**21
Quill (first space-based SAR), **18.**6
ScanSAR, **18.**12 to **18.**13, **18.**24
Seasat, **18.**7 to **18.**8
Shuttle Imaging Radar (SIR), **18.**8
spotSAR **18.**23
strip map, **18.**23
transmitters, **18.**19

Space-time adaptive processing (STAP), **3.**23 to **3.**31, **3.**33, **5.**2 to **5.**3, **5.**17, **24.**21 to **24.**22
in HF OTH radar, **20.**53, **24.**52
Spark-gap transmitter, **10.**2
SPASUR, **23.**2, **23.**9, **23.**11 to **23.**13
Speckle, **16.**55
Spectral characteristics of clutter, **2.**11 to **2.**16
Spectral noise in doppler radars, **10.**20 to **10.**21
Spectrometers, **16.**20, **16.**26 to **16.**28
Spectrum, of pulse doppler radar, **4.**4 to **4.**6
Specular point, **14.**20 to **14.**21
Specular scatterers, **14.**3
Sphere, radar cross section of, **14.**5 to **14.**6
Spherical reflector antenna, **12.**24 to **12.**25
Spotlight SAR, **17.**4 to **17.**5
SpotSAR, **18.**12, **18.**23
Spread-F region, **20.**19
Spurious output, in transmitters, **10.**19
Spurious response of mixers, **6.**11 to **6.**13
SR-71 low cross section aircraft, **14.**39 to **14.**41
Stability requirements in pulse doppler, **4.**27 to **4.**31
Stagger PRF design for MTI, **2.**39 to **2.**46
STALO, **6.**3, **6.**14 to **6.**20
Standard atmosphere, **26.**3
Stand-off jamming (SOJ), **24.**43 to **24.**44
Station keeping, airborne, **5.**27
Stealth, **14.**2, **14.**36, **24.**38
Stratiform rain, **19.**26
Stretch pulse compression, **6.**10, **6.**22, **8.**31 to **8.**36
in SBR altimeter, **18.**40 to **18.**41
String, radar cross section of, **14.**6 to **14.**7
Strip map, **18.**23
Stripmap SAR (or “strip” SAR), **17.**3 to **17.**4
Subarray adaptivity to ECM, **24.**25 to **24.**27
Subarrays, **13.**43 to **13.**44, **13.**49 to **13.**50
Subclutter visibility, in MTI, **2.**22
Subrefraction, **26.**6
Sudden ionospheric disturbance, **20.**18
Sunspot number, **20.**17
Superclutter visibility, **2.**84
Superheterodyne, **6.**1
Superrefraction, **26.**7
Superresolution, **24.**30 to **24.**31
Surface discontinuities, and radar cross section, **14.**3
Surface ducts, **26.**10 to **26.**11
Surface traveling-wave echoes, **14.**36, **14.**38
Surface ace-wave absorbers, **14.**38
Surface winds, and HF radar, **20.**37 to **20.**38
Surveillance radar, **1.**5
Surveillance radar equation, **1.**12
Synthesizer, digital, **25.**20 to **25.**22
Synthetic aperture radar (SAR)
autofocus, **17.**15
basic principle of **17.**1
burst mode, **18.**23 to **18.**24
clutterlock, **17.**15
comparison with optical imaging, **17.**18 to **17.**21
comparison with real-aperture radar, **17.**9 to **17.**10
crossrange resolution, **17.**7 to **17.**8
design issues for SBR, **18.**16 to **18.**24
doppler beam sharpening, **5.**34 to **5.**36, **5.**37, **17.**2, **17.**3
early history of, **17.**2
and ECCM, **24.**48 to **24.**51
fast-time processing, **17.**6
focused, **17.**3 to **17.**5
foliage penetration, **17.**33 to **17.**34

and ground penetrating radar, **21.**4, **21.**13
height measurement with, **17.**27 to **17.**23
image quality, **17.**16 to **17.**21
interferometric (InSAR), **17.**5, **17.**23 to **17.**24, **17.**30 to **17.**33
inverse (ISAR), **5.**23 to **5.**24, **5.**31 to **5.**33, **9.**37, **17.**5, **24.**51 to **24.**52
Joint STARS, **17.**23 to **17.**24
key aspects of, **17.**10 to **17.**15
key equations, **17.**21 to **17.**22
motion compensation in, **17.**12
moving targets in, **17.**23 to **17.**27
multiplicative noise ratio, **17.**17 to **17.**18
point-spread function, **17.**16
polarimetric, **17.**22
pulse repetition frequency requirements for, **17.**13 to **17.**15
range migration, **17.**15
range resolution, **17.**6 to **17.**7
range and velocity contours, **17.**10 to **17.**12

resolution, **17.**5 to **17.**10
resolution (crossrange) examples, **17.**5 to **17.**6
shadows in, **17.**28
signal-to-noise ratio, **17.**16 to **17.**17
slow-time processing, **17.**6
for space applications, **18.**5 to **18.**29
specific aspects of, **17.**22 to **17.**34
spotlight, **17.**4 to **17.**5
squinted stripmap, **17.**4
stereo, **17.**29 to **17.**30
stripmap, **17.**3 to **17.**4, **17.**14 to **17.**15
types of, **17.**2 to **17.**6
vibrating targets in, **17.**25 to **17.**27
unfocused, **17.**3
Washington monument image, **17.**18 to **17.**21
System instabilities and MTI, **2.**65 to **2.**73
System noise, **6.**5

T

TACCAR, **3.**4 to **3.**9, **3.**33
Target acquisition, in tracking radar, **9.**20 to **9.**21
Target caused errors in tracking, **9.**26 to **9.**37
Target cross section. *See* radar cross section
Target noise, in tracking radar, **9.**26 to **9.**37
Target recognition, **1.**7
cooperative, **5.**22 to **5.**23
noncooperative **5.**22 to **5.**24
Target resolution, in automatic detection, **7.**19 to **7.**20
Taylor weighting or illumination, **8.**7 to **8.**9, **13.**29
TecSAR, **18.**12
TEM horn, **21.**27 to **21.**30
Terminal Doppler Weather Radar (TDWR), **19.**1, **19.**30 to **19.**31
MTI filter design for, **2.**47 to **2.**50
Terrain avoidance, **5.**28 to **5.**29
Terrain-bounce jamming, **24.**43
Terrain database and radar, **5.**30
Terrain following, **5.**28 to **5.**29
Terrain height estimation, **5.**29 to **5.**30
Terrain scattered interference, **24.**43
TerraS AR-X, **18.**12
Test ranges, indoor, **14.**32 to **14.**35
Test ranges, outdoor, **14.**30 to **14.**32
THAAD radar, **13.**68
Thunderstorm prediction, **19.**32
Time-delay scanning, **13.**7
Time-frequency coded waveforms, **8.**25 to **8.**26
Topex altimeter parameters, **18.**35
Tornado detection, **19.**28 to **19.**29
Towed decoy, **24.**8
T/R module, **10.**28
Track association, **7.**38 to **7.**41
Track file, **7.**23 to **7.**25
Tracking, automatic, **7.**22 to **7.**46
Tracking radar, **1.**6
acquisition with, **9.**20 to **9.**21
automatic gain control (AGC), **9.**33 to **9.**35
conical scan, **9.**16 to **9.**17
dual band, **9.**24 to **9.**25
and ECCM, **24.**43
error reduction techniques, **9.**46 to **9.**47
errors in, **9.**26 to **9.**46
external causes of error, **9.**37 to **9.**42
glint in, **9.**30 to **9.**35, **9.**47

limitations in performance, **9.**44 to **9.**45
monopulse, **9.**3 to **9.**16
two-channel, **9.**14 to **9.**15
multipath error reduction, **9.**46 to **9.**47
*n*th-time-around tracking, **9.**24
on-axis tracking, **9.**25 to **9.**26
range tracking in, **9.**21 to **9.**24
sequential lobing, **9.**16 to **9.**17
servosystems for, **9.**17 to **9.**19
sources of error in, **9.**26
updating tracks, **7.**26 to **7.**30
TRAKX, dual band monopulse radar, **9.**24
Transmit/receive module characteristics, **11.**29 to **11.**31
Transmit signal digital processing, **25.**20 to **25.**22
Transmitter-related ECCM, **24.**31 to **23.**32
Transmitters, **1.**2, **1.**4
clustered-cavity klystrons, **10.**12 to **10.**13
constant efficiency amplifier, **10.**21 to **10.**22, **10.**26
crossed-field amplifiers (CFA), **10.**16 to **10.**17
extended interaction klystron (EIK), **10.**11
grid-controlled tubes, **10.**21 to **10.**23, **10.**25
gyrotrons, **10.**17 to **10.**19, **10.**26
for HF OTH radar, **20.**23 to **20.**26
linear-beam amplifiers, **10.**4 to **10.**13

Transmitters (*cont.*)
klystrons, **10.**5 to **10.**8, **10.**26
multiple-beam, **10.**7 to **10.**8
magnetrons, **10.**14 to **10.**16, **10.**26
microwave power module (MPM), **10.**13
modulators for, **10.**23 to **10.**24
MTBF of tubes, **10.**6 to **10.**7, **10.**10, **10.**15, **10.**16
oscillator vs. amplifier, **10.**4
role in radar, **10.**1 to **10.**2
solid-state, **10.**26 to **10.**28.
See also solid-state
spectrum control, **10.**19 to **10.**21
traveling wave tube, **10.**3, **10.**8 to **10.**10, **10.**26
Twystron, **10.**11
types of, **10.**2 to **10.**4
variants of klystrons and TWTs, **10.**11 to **10.**13
which type to use, **10.**25 to **10.**28
Transponder, **24.**6
Trapping, in ducting, **26.**7
Tropical Rainfall Measurement Mission (TRMM) SBR, **18.**62, **19.**39
Troposphere, **26.**2
effect on tracking accuracy, **9.**41 to **9.**42

U
UHF, **1.**15
Ultralow sidelobes, **24.**37
Ultrawideband (UWB), **21.**1
Updating tracks, **7.**26 to **7.**30

V
Velocity ambiguities, in meteorological radar, **19.**12 to **19.**13
Velocity-Azimuth-Display (VAD), **19.**31 to **19.**32
Velocity estimation, **19.**22
Velocity of propagation, in GPR, **21**.11
Velocity search, in pulse doppler, **4.**36
Venera, **18.**43 to **18.**45
Vessel tracking service (VTS) radars, **22.**1
Vessel tracking services, **22.**29 to **22.**31
VHF, **1.**15
Vibrating targets, in SAR, **17.**25 to **17.**27
Volume Search Radar (VSR), **13.**62 to **13.**63

W
Wake detection, **5.**31
Warloc millimeter-weave radar, **10.**18
Washington Monument SAR image, **17.**18 to **17.**21
Water vapor, attenuation in, **19.**7
Water vapor measurement with radar, **19.**32 to **19.**33
Wave spectrum, of the ocean, **15.**3 to **15.**5
Waveform upconversion, **6.**50
Waveguide models, for propagation, **26.**15 to **26.**26
Weather avoidance, with fighter radar, **5.**24
Weather radar. *See* meteorological radar
Wedgetail AEW radar, **3.**3
Wide bandgap semiconductors, **11.**15 to **11.**17
Wind measurement, **19.**31 to **19.**32
Wind profiler, **19.**39 to **19.**40
WSR-88D (Nexrad), **19.**1, **19.**16 to **19.**17

X
X-45C low cross section unmanned combat aircraft, **14.**42 to **14.**43
X band, **1.**17
X-Band Radar (XBR), **13.**68 to **13.**69

Z
Z (radar reflectivity factor), **19.**4
Z_e (effective reflectivity factor), **19.**6